NATURE | EDITORIAL

Redirection home

Europe's researchers should grab every opportunity to ensure that funds redirected towards strategic investment will not miss science altogether.

04 February 2015

The almost 19,000 followers of @EU_H2020 — the official Twitter feed of the European Union's flagship funding scheme Horizon 2020 — have already had much to discuss this year. Highlights include plans being drawn up by research commissioner Carlos Moedas on how to manage scientific advice after the abrupt axing of the chief adviser post held by Anne Glover; a live stream of a green-transport event; and the announcement of the first grants, including cash for projects to work on robots that wash floors and harvest sweet peppers.

However, @EU_H2020 has been quiet on a move by commission president Jean-Claude Juncker to raid the Horizon 2020 budget for money to help set up a continent-wide investment fund. The floor-washing robots are safe: Juncker wants to drain the cash — some €2.7 billion (US$3.1 billion) — from other parts of the budget, details of which were announced through more traditional routes last month.

Hardest hit is the European Institute of Innovation & Technology in Budapest, which will lose €350 million over the next six years. The European Research Council will lose €221 million, starting next year.

Related stories

- Economic divide taking toll on European science
- Unpaid bills complicate EU science budget crisis
- After the Berlin Wall: Central Europe up close

Also targeted is cash earmarked for projects across the continent over the coming years, including from information and communications technology, which will lose €307 million, food (€181 million) and nanotechnology, biotechnology and other advanced manufacturing techniques (€169 million).

If Juncker's proposal is approved by the European Parliament and Council, then the €2.7 billion will form part of a €16-billion European Fund for Strategic Investments that the European Commission hopes will stimulate state and private investment and lift the continent's stagnant economy.

@EU_H2020 might have been quiet on the move, but there have been howls of protest from those on the receiving end of the cuts.

"Horizon 2020 is not a lemon! Stop squeezing it!" was the sharp response from the League of European Research Universities in Leuven, Belgium, when the cuts were first suggested last year. And the advocacy group EuroScience said that it "is not in principle against using a small part of the Horizon 2020 budget for this purpose", but that taking the money from the European Research Council sent "a very bad signal". The European Research Area's Stakeholders Platform, an umbrella group of various

organizations expressed "great concern" and warned that the cuts would undermine research and innovation efforts across Europe.

> "Scientists must lobby for research and innovation to have a central role."

In response, European Commission officials say that the cuts come from an already generous budget — the original €80 billion in spending planned through Horizon 2020 makes it the most lucrative research funding scheme of its type in the world. The €2.7-billion reduction, they point out, could have been worse, and leaves the bulk of the programme intact. They argue that the funds will not truly be lost from science and research; they will return with interest when the strategic fund begins to bear economic fruit.

Perhaps. But it is easy to have sympathy for the organizations that were banking on that money and must now try to fill the hole. It is also easy to question the use of the word 'strategic' in the title of the fund. Strategy is long-term, and the best and most enduring route to prosperity must remain the careful allocation of investment to research on science and technology — both pure and applied.

Still, as *Nature* has argued before, scientists must accept that the boom times are over, at least for now. Money is tight and priorities are shifting. Those in Europe would do well to remember that.

The new fund could be up and running as soon as September, so some scientists could still be waiting to hear whether they will join the pepper-picking robot researchers in receiving a Horizon 2020 grant (chances are, they won't, the programme is massively oversubscribed, sorry). In principle, research could yet benefit from the redirected money, but scientists and their representatives must lobby for research and innovation to have a central role in the projects — infrastructure and the rest — in which the new fund will invest. The European Research Area's Stakeholders Platform has suggested amendments to the proposed legislation to make that happen, including giving researchers a say in how the money is allocated, and European officials should listen to that advice.

Science may have lost out on the money, but it should not miss out on the opportunity.

Related stories and links

From nature.com

- **Economic divide taking toll on European science**
  07 January 2015
- **Unpaid bills complicate EU science budget crisis**
  26 November 2014
- **After the Berlin Wall: Central Europe up close**
  05 November 2014

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments   **Subscribe to comments**

There are currently no comments.

See other News & Comment articles from *Nature*

Nature   ISSN 0028-0836   EISSN 1476-4687

House of cards

Western institutions must speak out against human-rights abuses in their partner countries.

03 February 2015

When the leaders of many of the world's democracies flocked to Saudi Arabia last week to offer their condolences on the death of King Abdullah, many critics called it hypocrisy. They did so, too, when Saudi officials marched in Paris two weeks earlier to defend freedom of expression following the terrorist attacks there.

After all, Saudi Arabia comes near the bottom of the world league in terms of freedoms, such as the right to dissent, to freedom of expression or to practise any religion other than Islam, and has a track record of brutal human-rights abuses and political and religious oppression. But the kingdom's oil and strategic geopolitical importance in the turbulent Middle East means that it has long enjoyed strong ties with the West.

Some scientists have been drawn to the desert state too, not least to the King Abdullah University of Science and Technology (KAUST) in Thuwal, a graduate university created by the king in 2009, which has a US$20-billion endowment. The university is the flagship of Abdullah's efforts both to build a knowledge-based society in a country with little science base and to help distance science and education from the stifling influence and control of conservative clerics.

Related stories

- Saudi university backs slow road to modernization
- Scientists protest detention of Palestinian physicist
- Task force seeks reform at Muslim-world universities

More related stories

As we report on page 18, some of these scientists have become caught up in the controversy over Saudi Arabia's human-rights record. An international outcry has been sparked by the Saudi authorities' flogging of the activist Raif Badawi in a public square in January — the first 50 of a sentence of 1,000 lashes, along with 10 years in prison, for posts that he introduced on his website for social and political discussion.

The Badawi case once again highlights the responsibility of researchers and scientific institutions who collaborate with authoritarian and repressive regimes such as Saudi Arabia to denounce human-rights abuses. Eighteen Nobel laureates explicitly raised that point in a letter last month to the president of KAUST, calling for "influential voices in KAUST" to speak out against Badawi's brutal treatment, arguing that no university can be viable in a society lacking basic freedoms.

Some scientists and their institutions, such as the US National Academies of Science, have a long history of speaking out to defend freedoms, and of campaigning on behalf of persecuted academics and

activists, although too many others remain silent. Still, there are concerns that such lobbying has lessened in recent years, with several scientific human-rights bodies, including those of the New York Academy of Sciences and the American Association for the Advancement of Science, shifting their focus to scientific diplomacy and softer human-rights issues, such as access to education, clean water, food and health care. Some have argued that working to open up repressive countries is more effective in the long term than publicly embarrassing them over individual cases of abuse.

> "Change cannot be expected to come quickly in Saudi Arabia."

Others have rightly expressed concern that scientists and their institutions may be increasingly reluctant to speak out to avoid jeopardizing collaborations with countries, including China, that have dismal human-rights records. The many Western universities that have partnerships with KAUST and other Saudi institutions benefit from petrodollars, and the leading researchers who have joined the KAUST faculty benefit from competitive salaries and state-of the-art laboratory conditions. Western universities have also gained from the influx of hundreds of thousands of fee-paying Saudi students under a generous scholarship scheme established by King Abdullah.

What can scientists there achieve by speaking out? Foreign researchers working at KAUST who were contacted by Nature seem sincerely convinced that, by educating and broadening the horizons of young Saudi Arabians, they can do more good by working to help to slowly open up the regime. The scientists are to be applauded for their efforts — this journal has long backed scientific cooperation as a form of diplomacy, for example with Iran, and has similarly opposed proposed scientific boycotts of Israel.

Unfortunately, change cannot be expected to come quickly in Saudi Arabia because of the unique complexity of its society and culture. As Europe's Enlightenment was taking shape in the eighteenth century, pushing back against religious authority and ushering in modern science, the Arabian peninsula was heading in the opposite direction. The Saudi state was born at the time out of an unholy alliance between Ibn Saud, a tribal leader, and Muhammad Ibn Abd al-Wahhab, the leader of Wahhabism, an extreme fundamentalist sect of Sunni Islam. That pact shapes Saudi rule and society to this day, resulting in a symbiotic agreement, with the conservative clerics giving the monarchy its support in return for their power to impose a society based on radical Islam, and an extreme form of sharia law.

But there does not need to be a conflict between defending individual cases — either publicly or by more diplomatic, behind-the-scenes pressure — and broader outreach efforts. We need both. Campaigns for persecuted individuals whose plights otherwise risk going unnoticed can also, as in Badawi's case, send the powerful message that the world is watching. Scientists at KAUST are perhaps not best placed to speak out, being at risk of potential retribution. But Saudi Arabia benefits hugely, not least in terms of its international image, from prominent collaborations with Western research organizations and universities, which have a duty to use that leverage to speak out on abuses, and to call for greater democratic reforms — both publicly and in their private dealings with their Saudi partners.

Related stories and links

From nature.com

- **Saudi university backs slow road to modernization**
  03 February 2015
- **Scientists protest detention of Palestinian physicist**
  21 January 2015
- **Task force seeks reform at Muslim-world universities**
  20 January 2015
- **Changing attitudes in Saudi Arabia**
  21 March 2014
- **The case of Dr Hicheur**
  12 October 2011
- **Human-rights shift under fire**
  27 July 2011
- **Time to connect**
  05 March 2008
- **Scientists suffer human-rights abuses**
  25 October 2006
- **Forgotten plights**
  11 October 2006
- *Nature* **special: Islam and science**
- **Web Focus: AIDS medics in Libya**

From elsewhere

- **King Abdullah University of Science and Technology (KAUST)**
- **Copy of Nobel letter**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments   Subscribe to comments

There are currently no comments.

See other News & Comment articles from *Nature*

Nature   ISSN 0028-0836   EISSN 1476-4687

Thuwal, a graduate university created by the king in 2009, which has a US$20-billion endowment. The university is the flagship of Abdullah's efforts both to build a knowledge-based society in a country with little science base and to help distance science and education from the stifling influence and control of conservative clerics.

As we report on page 18, some of these scientists have become caught up in the controversy over Saudi Arabia's human-rights record. An international outcry has been sparked by the Saudi authorities' flogging of the activist Raif Badawi in a public square in January — the first 50 of a sentence of 1,000 lashes, along with 10 years in prison, for posts that he introduced on his website for social and political discussion.

The Badawi case once again highlights the responsibility of researchers and scientific institutions who collaborate with authoritarian and repressive regimes such as Saudi Arabia to denounce human-rights abuses. Eighteen Nobel laureates explicitly raised that point in a letter last month to the president of KAUST, calling for "influential voices in KAUST" to speak out against Badawi's brutal treatment, arguing that no university can be viable in a society lacking basic freedoms.

Some scientists and their institutions, such as the US National Academies of Science, have a long history of speaking out to defend freedoms, and of campaigning on behalf of persecuted academics and activists, although too many others remain silent. Still, there are concerns that such lobbying has lessened in recent years, with several scientific human-rights bodies, including those of the New York Academy of Sciences and the American Association for the Advancement of Science, shifting their focus to scientific diplomacy and softer human-rights issues, such as access to education, clean water, food and health care. Some have argued that working to open up repressive countries is more effective in the long term than publicly embarrassing them over individual cases of abuse.

Others have rightly expressed concern that scientists and their institutions may be increasingly reluctant to speak out to avoid jeopardizing collaborations with countries, including China, that have dismal human-rights records. The many Western universities that have partnerships with KAUST and other Saudi institutions benefit from petrodollars, and the leading researchers who have joined the KAUST faculty benefit from competitive salaries and state-of the-art laboratory conditions. Western universities have also gained from the influx of hundreds of thousands of fee-paying Saudi students under a generous scholarship scheme established by King Abdullah.

What can scientists there achieve by speaking out? Foreign researchers working at KAUST who were contacted by *Nature* seem sincerely convinced that, by educating and broadening the horizons of young Saudi Arabians, they can do more good by working to help to slowly open up the regime. The scientists are to be applauded for their efforts — this journal has long backed scientific cooperation as a form of diplomacy, for example with Iran, and has similarly opposed proposed scientific boycotts of Israel.

> *"Change cannot be expected to come quickly in Saudi Arabia."*

Unfortunately, change cannot be expected to come quickly in Saudi Arabia because of the unique complexity of its society and culture. As Europe's Enlightenment was taking shape in the eighteenth century, pushing back against religious authority and ushering in modern science, the Arabian peninsula was heading in the opposite direction. The Saudi state was born at the time out of an unholy alliance between Ibn Saud, a tribal leader, and Muhammad Ibn Abd al-Wahhab, the leader of Wahhabism, an extreme fundamentalist sect of Sunni Islam. That pact shapes Saudi rule and society to this day, resulting in a symbiotic agreement, with the conservative clerics giving the monarchy its support in return for their power to impose a society based on radical Islam, and an extreme form of sharia law.

But there does not need to be a conflict between defending individual cases — either publicly or by more diplomatic, behind-the-scenes pressure — and broader outreach efforts. We need both. Campaigns for persecuted individuals whose plights otherwise risk going unnoticed can also, as in Badawi's case, send the powerful message that the world is watching. Scientists at KAUST are perhaps not best placed to speak out, being at risk of potential retribution. But Saudi Arabia benefits hugely, not least in terms of its international image, from prominent collaborations with Western research organizations and universities, which have a duty to use that leverage to speak out on abuses, and to call for greater democratic reforms — both publicly and in their private dealings with their Saudi partners. ∎

Road test

Realizing the benefits of driverless cars will require governments to embrace the technology.

The government funding agency Innovate UK has launched a £10-million (US$15-million) project to study how autonomous, self-driving vehicles will fit into daily life in four parts of England: Greenwich, Coventry, Milton Keynes and Bristol.

Good job. That is the right kind of question to ask about driverless cars. As described in a News Feature on page 20, developers such as Google are making rapid progress on the vehicles. From a technical standpoint, the cars could be ready for widespread deployment within a decade. But when and how they will hit the streets depends on how well people accept and trust them.

Consider, for example, the obvious economic question: will people be able to afford them? Thanks to the need for sophisticated equipment, the vehicles are likely to be much more expensive than their conventional counterparts, at least initially. And that means that buyers will need to see correspondingly large benefits.

A frequently cited benefit is safety: advocates insist that the vehicles could all but eliminate accidents. But convincing people that driverless cars can do away with human accidents and not make robot-minded mistakes of their own is likely to take a good number of years and millions of kilometres of accident-free test drives.

And when accidents do happen — as they surely will — public reaction will depend on the specifics of the event, and those are hard to predict. The legal issues may be even tougher. Right now, equipment failures are rare and the responsibility almost always rests with a driver. But with driverless vehicles, the courts and insurance companies will have to figure out how to apportion liability among the vehicle's occupants (who may be dozing off), the car maker, the software developers and even the mapping algorithm.

Another much-touted benefit is fuel efficiency. But that is unlikely to be realized until most cars are equipped with systems that allow them to communicate with one another (called V2V systems) and with traffic signals to minimize stop-and-go traffic.

Of course, some wealthy people will doubtless take the plunge. But the most important early adopters will probably be fleet operators: driverless ride-share systems could function as a new form of mass transit. And if the door-to-door service encourages more people to give up their car, then some of the vast areas devoted to parking could be put to other uses.

Governments are likely to be crucial to the transition — not least because many of the benefits accrue to society as a whole. A good example is being set by the United States, which is considering a mandate that would greatly speed up the transition by requiring V2V radios in every new US car. Other countries should follow suit. To make such moves fully effective, however, local governments will need to start upgrading roadways with smart signals designed to optimize traffic flow — assuming they can find the money. ∎

↻ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunqv

JEREMY SUTTON-HIBBERT

Major biodiversity initiative needs support

An effort aimed at protecting ecosystems, modelled on the agency battling climate change, will need protecting from powerful enemies, warns **Ehsan Masood**.

There was something different about Rajendra Pachauri, chairman of the Intergovernmental Panel on Climate Change (IPCC), when he rose to address a major conference on biodiversity in Bonn, Germany, late last month. His signature green tie was absent; a red alternative hung in its place.

Red for danger, Pachauri said, to acknowledge the peril facing ecosystems and much of the natural world. Danger, he added — pausing for effect — was not a word he could use in the highly politicized context of climate change. Researchers who investigate and log Earth's diminishing biodiversity, he was hinting, have yet to encounter the kind of distortions and politicization that are a regular feature for those who work on global warming. But for how long will that continue?

The Bonn conference was the third plenary meeting for a major initiative that explicitly aims to mimic the workings and impact of the IPCC, including eventually drawing up laws that would put a scientific brake on rampant development. As such, it is likely to make powerful enemies. One of its first reports will assess the state of pollinating insects. Others will explore the highly charged question of how to value ecology. The red tie is a sign of things to come.

The initiative — the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) — was set up three years ago, although the idea was first mooted in the 1990s. The mood in Bonn was upbeat as delegates agreed its annual US$9-million budget and put the seal on a busy programme of work for the next five years.

Recent controversies over the IPCC — claimed errors in its reports and debate about whether the panel should even continue in its present form — might seem to make the organization a dubious role model. Does the world really need another lumbering process that involves hundreds of scientists, who anyway need to have their final work signed off by representatives of politicians?

Such a view underestimates the IPCC's impact in one crucial area: to provide political impetus and an evidence-backed mandate for international legislation. The agency's second assessment report, the one confirming a human fingerprint on climate, overcame political dissent at the time and led directly to the 1997 Kyoto Protocol, which remains the only agreement that legally binds states to reduce their greenhouse-gas emissions. The protocol is unfashionable in some climate-policy circles now, but the IPCC remains the model to drive a majority of the world's governments to change laws in response to scientific findings.

The founders of the IPBES want something equivalent to protect biodiversity. The need is urgent. The existing United Nations biodiversity convention, an international agreement in which member countries promise to protect, sustainably use and share the benefits of biodiversity, lacks teeth and has made little impact on slowing biodiversity loss. One-eighth of birds, one-quarter of mammals and one-third of amphibians are understood to be facing the threat of extinction, IPBES chair Zakri Abdul Hamid told the conference. The present consensus is that the rate of extinction is somewhere between 100 and 1,000 times the pre-industrial background rate.

The first global report from the new biodiversity panel — similar to the periodic landmark assessment reports from the IPCC — is due in 2019. The initial shots in the conflict that could follow have already been fired. The United States so far finds itself unable to pay for its scientists to contribute; most of the money for the exercise, moreover, is coming from European countries. Not coincidentally, the United States has still to ratify the biodiversity convention, which many lawmakers, Republicans in particular, regard as anti-growth.

Still, insiders expect the US government and its national institutions to play a bigger part in the coming months. One of its tasks, along with Europe and Latin America, will be to protect the role that conservation and industry groups have in the IPBES as observers. Some countries, notably China, seem to want to restrict this.

There is one major difference from the IPCC. Each IPBES assessment must include reference citations to indigenous knowledge, and every review panel must include experts in this. That is partly a concession to some developing countries that, for many years, resisted the idea of the IPBES, fearing that it would be based, like most IPCC reports, on studies in peer-reviewed European-language journals. It also reflects the fact that, by definition, most of the planet's remaining biodiversity is in developing nations.

In the science ministries of powerhouse nations, the study of indigenous knowledge is viewed as soft, flaky even. Compared with fields such as plant genetics, it is also less likely to be recognized by many leading science academies. Already, some representatives from Europe have complained that they cannot find suitably qualified individuals to conduct or review assessments.

Anticipating this, the IPBES has set aside funding to train and identify suitable experts, especially from developing countries. IPBES leaders should cast the net further, and draw in more experts from the social sciences and others — through learned societies of humanities scholars, for example. Biodiversity needs all the help it can get. ∎

BY DEFINITION, MOST OF THE PLANET'S REMAINING BIODIVERSITY IS IN DEVELOPING NATIONS.

Ehsan Masood *is editor of* Research Fortnight *and a co-author with Daniel Schaffer of* Dry: Life Without Water.
e-mail: ehsan.masood@researchresearch.com; Twitter: @EhsanMasood

What drives sugar addiction

Researchers have identified the brain circuits that compel mice to seek out sweet treats.

Kay Tye of the Massachusetts Institute of Technology in Cambridge and her colleagues genetically engineered mice so that the neurons in a brain circuit involved in reward processing would fire when exposed to light. When the researchers activated these neurons, the animals sought sugar more frequently through a port in their cage, even when they received a mild electric shock to their feet in the process. Switching the neurons off stopped the sugar-seeking behaviour, but did not prevent the mouse from eating its normal food.
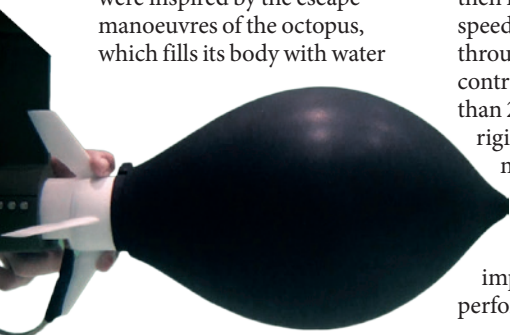
The researchers propose that targeting this pathway could be a possible therapy for compulsive overeating.
Cell **160,** 528–541 (2015)

Robot zips away like an octopus

A self-propelling, octopus-inspired rocket can zoom through water using energy more efficiently than fish that accelerate rapidly to escape.

Gabriel Weymouth of the University of Southampton, UK, and his colleagues were inspired by the escape manoeuvres of the octopus, which fills its body with water

KATHRYN LANGIN

Birds diversify at close quarters

A population of birds in California has evolved diverse bill sizes and shapes, even within a small geographic area.

Diversification within a species is thought to occur mainly when populations are separated by geographical barriers, such as mountains or bodies of water. To see if this happens in a single population, Kathryn Langin at Colorado State University in Fort Collins and her colleagues studied more than 500 island scrub-jays (*Aphelocoma insularis*; pictured), which have small ranges and live only on the 250-square-kilometre Santa Cruz Island in California.

The team found that birds living in pine forests had longer, shallower bills — presumably for prying open pine cones — than jays in oak forests, even though the two habitats were next to each other. Adaptations at this microgeographic level could be more common than once thought, even for mobile animals, the authors suggest.
Evolution http://doi.org/zt3 (2015)

and then quickly expels it to dart away. On the basis of this principle, the authors built their rocket using a flexible hull, which they inflated with water (**pictured**). They then measured the rocket's speed as it shot the fluid out through its base. As the rocket contracted, it achieved more than 2.6 times the thrust of a rigid rocket doing the same manoeuvre.

The researchers calculate that making the robot bigger would improve its accelerating performance, and suggest that their technology could be used in underwater vehicles.
Bioinspir. Biomim. http://dx.doi.org/10.1088/1748-3190/10/1/016016 (2015)

Smoke makes tornadoes worse

Smoke in the air could increase the likelihood of tornadoes forming, as well as their intensity.

Pablo Saide of the University of Iowa in Iowa City and his colleagues studied a major tornado outbreak that killed more than 300 people in April 2011 in the southeastern United States. It occurred when smoke particles from fires in Central America drifted across the region, which the team's calculations suggest led to atmospheric changes, such as thicker and lower clouds. These had several cascading effects, including stronger wind shear, that make tornadoes more likely to form and more severe in areas that are prone to such storms.

Meteorologists may need to

GABRIEL D. WEYMOUTH AND VIGNESH SUBRAMANIAM

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Mathematical model helps scientists decide where to submit their papers

A study that ranks ecology journals on the basis of citation numbers and review times attracts attention online.

Chris Woolston

29 January 2015

Deciding where to submit a scientific manuscript is a strategic decision for researchers who are balancing the desire for high impact with the need for quick publication. Two ecologists have tried to crack that conundrum by creating mathematical models that quantify the potential value of submitting to any of 61 ecology journals. In a study in *PLoS ONE*[1], they conclude that *Ecology Letters*, *Ecological Monographs* and *PLoS ONE* are good choices for researchers seeking a large number of citations but relatively fast review times. Researchers on social media appreciated the effort. Ross Mounce, an evolutionary biologist at the University of Bath, UK, tweeted:

Ross Mounce
@rmounce

Where Should I Send It? Optimizing the Submission Decision Process journals.plos.org/plosone/articl … HT @Protohedgehog @LaurenMaggio Super interesting!

RETWEETS　　　FAVORITES
2　　　　　　　2

8:04 AM - 26 Jan 2015

But Marcel Holyoak, editor-in-chief of *Ecology Letters*, says that the analysis has limitations and is probably "most useful for a naive author without experience on which journals to target".

The *PLoS ONE* paper evaluated the 61 journals on the basis of data on acceptance rates and decision

times provided by the publications (*Nature* was not included in the analysis because the journal did not provide all the necessary data). One of the models is designed to maximize citations over a given time period, while taking into account the potential hassle and delay that comes with rejections and resubmissions. The model suggests that, in many instances, *PLoS ONE* would fall into a sweet spot that provides a moderate pay-off in citations and relatively little frustration. For example, submitting first to *PLoS ONE* instead of to a higher-impact journal such as *Ecology Letters* or *Science* might cost a paper up to 14 citations, but would probably avoid at least one revision while saving between 30 and 150 days of publication delay. The model also suggests that submitting to relatively low-impact journals, such as *Wildlife Monographs* or the *Russian Journal of Ecology* instead of *PLoS ONE*, would significantly reduce citations without much benefit in terms of waiting time or the chance of rejection.

The paper offers an important lesson, says Mounce: submitting to top-tier journals isn't always the best strategy, especially for scientists who need to rack up publications in a hurry for job promotions. "I know people who essentially had to quit academia because they chose the high-risk approach rather than just publishing their work in decent outlets," he says.

Jean-Michel Ané, a microbiologist and plant scientist at the University of Wisconsin–Madison, shared the article on the social publishing site scoop.it with the comment: "It would be great if they could do that for other disciplines like … microbiology and plant sciences for instance!"

Ané said in an interview that choosing where to submit a paper can be a high-stakes game when careers rest on it. Unfortunately, he says, some of the key variables in the decision can be hard to assess. "A journal may say that they review a paper within 10 days, but the reality can be very different," he says. "A database where those parameters could be compiled and updated regularly would be super useful."

Holyoak says that he's glad *Ecology Letters* fared well in the analysis, but he also believes that the model "is a bit crude", largely because it assumes that each article has the same quality, which he sees as an oversimplification. Santiago Salinas, an ecologist now at the University of the Pacific in Stockton, California, and co-author of the paper, acknowledges that the merits of a manuscript should play an important part in submission decisions. He adds that the model originally had a "paper quality" variable but it proved to be too subjective to be workable. "We tried to keep the model as flexible as possible," he says.

Many commenters on Twitter noted that the paper appeared in one of the journals that the model recommends. That's not an accident, Salinas says. "We actually used what we learned from the model to submit to *PLoS ONE*," he explains. "We followed our own advice."

References

1. Salinas, S. & Munch, S. B. *PLoS ONE* **10**, e0115451 (2015)

    Show context   Article PubMed

For the best commenting experience, please login or register as a user and agree to our Community
Guidelines. You will be re-directed back to this page where you will see comments updating in real-time
and have the ability to recommend comments to other users.

1 comment   Subscribe to comments

Ross Mounce` · 2015-01-29 06:13 PM

Here's the rest of my comments on this interesting paper:

http://rossmounce.co.uk/2015/01/29/manuscript-submission-modelling-my-comments-in-full/

See other News & Comment articles from *Nature*

partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

SEVEN DAYS
The news in brief

US-budget hope

US President Barack Obama's budget plan for 2016 has given scientists a ray of hope — although the proposals now face a rough road in Congress. The US$4-trillion budget proposal released on 2 February offers $146 billion for scientific research and development. This is a 6% rise in the pot of money split between civilian and defence programmes, and comes after years of austerity. The plan is expected to meet with strong opposition in the Republican-controlled Congress. See page 13 for more.

Vaccination boost

GAVI, an organization in Geneva, Switzerland, that supports vaccination in low-income countries, raised US$7.5 billion during a 27 January pledging conference in Berlin. The money, largely from Western governments, will pay for GAVI's immunization efforts between 2016 and 2020 and will augment $2 billion in previously committed funding. The extra cash should enable a further 300 million children to be vaccinated, preventing some 6 million deaths, the organization says.

China biosafety lab

China opened its first biosafety level-4 research laboratory on 31 January, enabling work on dangerous pathogens such as Ebola. The Wuhan National Biosafety Laboratory is funded and operated by the Chinese Academy of Sciences. The lab was approved in 2003 after the SARS virus outbreak revealed shortcomings in the country's ability to deal with emerging diseases. China enlisted
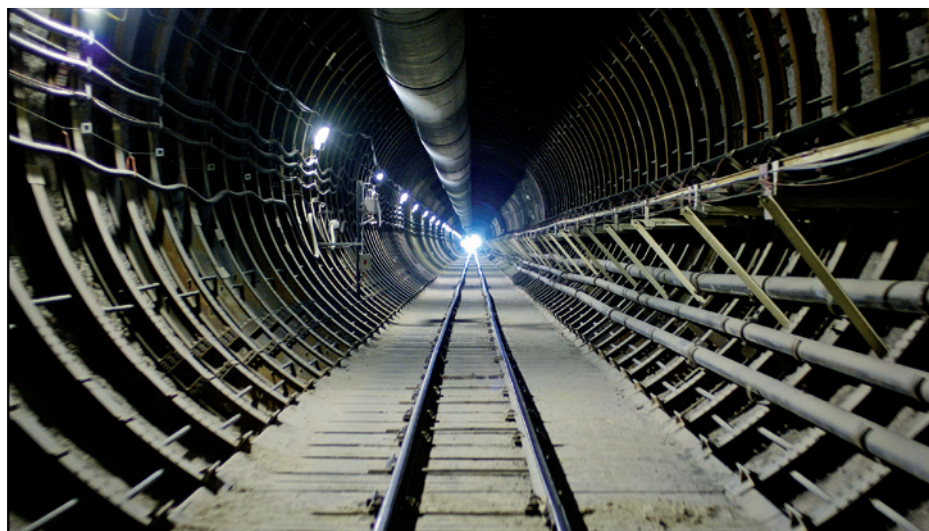
DAVID HOWELLS/CORBIS

Nuclear-waste plan is technically sound

The proposal to build a nuclear-waste repository at Yucca Mountain, Nevada, is technically sound, the US Nuclear Regulatory Commission in Rockville, Maryland, said on 29 January. The 2008 application by the Department of Energy was abandoned by the administration of US President Barack Obama in 2010, but a federal court ordered the commission to continue with the licensing process as long as it had the funds to do so. The commission notes, however, that construction would not have been possible — although a railway to bring in waste is already in place (pictured) — because the federal government had not secured land and water rights from Nevada, which opposes the project.

French assistance to design the facility, which it plans to use to develop diagnostics and vaccines against highly infectious diseases.

Dusty death

A signal thought to be the first evidence of gravitational waves was caused by dust in the Milky Way rather than being a relic of the Universe's first moments. The team that described the signal in March 2014 withdrew its claim on 30 January. Combined data from the BICEP2 telescope at the South Pole and the European spacecraft Planck revealed that the distinctive polarized light pattern spotted by the team was almost entirely due to Galactic noise. See page 16 for more.

Warmest year

The World Meteorological Organization (WMO) has officially ranked 2014 as the warmest year since modern temperature records began. According to an analysis of the three most widely used global climate data sets, which have been collected since around 1880, the mean global air temperature last year was 0.57 °C above the 14.00 °C average for the reference period 1961–90. Fourteen of the 15 warmest years on record have now occurred in the twenty-first century, the WMO notes. The three hottest years — 2014, 2010 and 2005 — are only a few hundredths of a degree apart, less than the margin of uncertainty of the measurements.

Ebola trials

Plans to test an experimental Ebola drug in Liberia have been scrapped, a US drug company said on 30 January. Chimerix, based in Durham, North Carolina, ended a trial of the antiviral brincidofovir because it failed to enrol enough people infected in the now-waning Ebola epidemic in West Africa. On 2 February, a different clinical trial of two experimental vaccines against Ebola virus began in Liberia. This trial aims to enrol around 27,000 people to determine whether the vaccines can prevent infection.
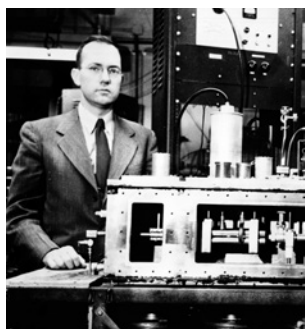
BUSINESS

Myriad settles suits

A long-running patent dispute over genetic testing is nearing its end. Myriad Genetics of Salt Lake City, Utah, confirmed last week that it has settled lawsuits against three other medical diagnostics firms over their use of genetic tests that analyse mutations in the *BRCA1* and *BRCA2* genes to estimate the risk of breast and ovarian cancer. The dispute began in 2009, and in 2013 the US Supreme Court struck down some of Myriad's patent claims to the tests, but the company argued that it was protected by other patents. Myriad is negotiating a settlement with the remaining four companies that it has sued over the tests.

PEOPLE

Laser laureate dies

Charles Townes (**pictured**), a US physicist who received a Nobel prize in 1964, died on 27 January, aged 99. While at Columbia University in New York in 1954, Townes built a device called a maser, which stimulates atoms to emit coherent bursts of microwave radiation — the principle that led to the first laser. After moving to the University of California, Berkeley, he used lasers to detect the first complex molecules in interstellar space and to find evidence for the black hole at the centre of the Milky Way.

Nobel chemist dies

French chemist Yves Chauvin, who shared the 2005 Nobel Prize in Chemistry, died on 27 January, aged 84. While working at the French Petroleum Institute in Rueil-Malmaison near Paris, Chauvin explained the mechanism behind a catalytic chemical reaction called metathesis, during which two molecules joined by doubly bonded carbon atoms swap their partners in a kind of dance. This understanding helped to develop catalysts to make metathesis more controllable, and it is now used widely in the synthesis of plastics, drugs and pesticides.

Pill creator dies

Chemist Carl Djerassi, widely known as the father of the birth-control pill, died on 30 January, aged 91. He emigrated to the United States to escape Nazi Germany's threat to his birthplace, Vienna, in 1939. Ten years later, he joined the pharmaceutical firm Syntex in Mexico City, where he led the research team that synthesized the first orally active steroid contraceptive, norethindrone. He later worked on biosynthesis and analytical chemistry at Stanford University in California, and wrote dozens of short stories, novels and plays.

Environment chief

The appointment of a scientist who has been dubbed gutsy and radical by the media as Communist-Party head of China's environmental protection ministry is raising hopes on combating pollution. Chen Jining, currently president of Beijing's prestigious Tsinghua University, was named as chief on 28 January and is expected to become environment minister in March. There are high hopes that he will improve the enforcement of China's environmental regulations, in particular a law that calls for tighter monitoring and punishment of polluters. China has often favoured economic growth over enforcing environmental laws, especially at the level of local government. See go.nature.com/n4wjjn for more.

PRIZES

Moon milestones

Five teams picked up Google Lunar X Prize 'milestone' prizes on 26 January. The awards, worth a total of US$5.25 million, were established in 2014 to recognize steps towards the prize's ultimate goal of landing a private spacecraft on the Moon by the end of 2016. The teams from Germany, India, Japan and the United States showed headway in landing, roving and imaging technology — albeit on Earth. Astrobotic, a spin-out company of Carnegie Mellon University in Pittsburgh, Pennsylvania, won prizes in all three areas. The awards were introduced after slow progress on the main $30-million challenge, the original deadline for which was in 2012.

↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

The US public generally supports science, but there seems to be a large gap between it and scientists on some controversial issues. Of about 2,000 adults surveyed, 79% say that science has made life easier for most people, found a poll by the American Association for the Advancement of Science (AAAS) and the Pew Research Center, a think tank in Washington DC. But researchers are left questioning the gulf between them and the public on certain topics. See go.nature.com/jnljfu for more.

OPINION GAP

On hotly debated scientific issues, scientists and the public differ greatly, reveals a poll by the AAAS and Pew Research Center.

Scientists ■ US public

Issue	Scientists	US public
Climate change is mostly due to human activity	87%	50%
It is safe to eat genetically modified foods	88%	37%
Humans have evolved over time	98%	65%
More nuclear power plants should be built	65%	45%
Favour use of animals in research	89%	47%
Childhood vaccines such as MMR should be required	86%	68%

NATURE | NEWS
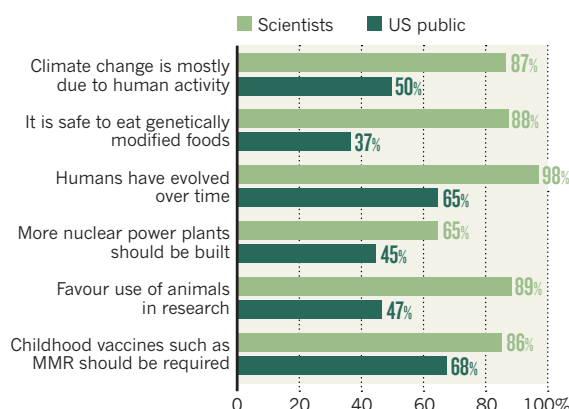
Obama budget seeks big boost for science

White House plan would increase research and development funding but faces rough road in Congress.

Boer Deng, **Richard Monastersky**, **Lauren Morello**, **Sara Reardon** & **Jeff Tollefson**

03 February 2015

J. Scott Applewhite/AP

US President Barack Obama's budget aims to raise spending on climate and biomedical research, among other areas.

When US President Barack Obama released his budget proposal on 2 February, he gave scientists and engineers a ray of hope — albeit one that is almost certain to be dimmed, if not extinguished.

Obama's US$4-trillion plan for fiscal year 2016 includes $146 billion for scientific research and development, a healthy 6% increase for a portfolio split roughly evenly between defence and civilian programmes. The proposal, which seeks to turn back years of fiscal austerity, is the opening salvo in what is likely to be a long war with the Republican-controlled Congress over government spending.

Top picks
from nature news

- Science pours in from Rosetta comet mission

"The problem, fundamentally, is that this budget is dead on arrival," says Michael Lubell, director of public affairs at the American Physical Society in Washington DC.

- GM microbes created that can't escape the lab
- Crunch time for pet theory on dark matter

But he says that Obama may yet succeed in convincing lawmakers to lift the spending caps known as sequestration, which were put in place in 2011. Congress agreed to ease the caps in 2014 and 2015, but a similar agreement for 2016 may be harder to broker. Republicans have expressed interest in ending limits on defence spending, and are seeking to compensate for that by making cuts to civilian programmes.

"We need a real budget, one that allows responsible investments in critical federal programs — including our national defense — without breaking the bank and pushing our country further into deficits and debt," said Hal Rogers (Republican, Kentucky), chairman of the House of Representatives appropriations committee, in a written statement.

In the meantime, the president is moving ahead with a budget request that aggressively lays out his priorities for the twilight of his term in office, which ends in January 2017. One top concern is climate change, an area of sharp disagreement between the White House and the Republican-led Congress. The multi-agency US Global Change Research Program would receive a 9% increase in 2016, to $2.7 billion.

Climate and energy

The US Environmental Protection Agency (EPA) would see an increase of roughly 6%, driving its budget to $8.6 billion — including $769 million for science and technology (see 'Budget highlights'). The agency would receive $239 million to carry out climate-change regulations and initiatives, and $25 million to help states to comply with a rule — expected to be finalized this year — that would limit greenhouse-gas emissions from power plants. The budget would also create a $4-billion fund to help states that want to enact even stricter emissions limits on the power sector.

Budget highlights

How science agencies fared in the budget (US$millions).

Agency	2014 actual	2015 estimated	2016 requested	Details
Biomedical research and public health				
National Institutes of Health	30,070	30,311	31,311	Increases spending on antimicrobial resistance by $100 million
Centers for Disease Control and Prevention	5,863	6,073	6,170	Includes $12 million to help developing countries to improve disease surveillance

Food and Drug Administration	2,561	2,596	2,744	Proposes a new agency to monitor food safety, currently a joint responsibility of FDA and agriculture department
Physical sciences				
National Science Foundation	7,131	7,344	7,724	Asks for a 5.2% rise from 2015
NASA (science)	5,148	5,245	5,289	Seeks $30 million to formulate a mission to the Jovian moon Europa
Department of Energy Office of Science	5,066	5,068	5,340	Allocates $2.7 billion to clean-energy and energy-efficiency programmes
National Institute of Standards and Technology	850	864	1,120	Nearly doubles funding for programmes to support manufacturing
Earth and environment				
Environmental Protection Agency	8,200	8,140	8,592	Allots $239 million to climate-change regulations and initiatives
National Oceanic and Atmospheric Administration	5,323	5,449	5,983	Allocates $2.4 billion to Earth-observing satellites.
US Geological Survey	1,032	1,045	1,195	Spends extra $32 million on science to increase climate resilience and adaptation

Source: White House Office of Management and Budget, US Public Law No. 113-235

Republicans are already promising to go after the EPA budget as they seek to head off the agency's new climate regulations. "This is a giant press release," says Frank Maisano, a lobbyist at Bracewell & Giuliani in Washington DC who represents energy-industry clients. "It's just the first marker in what is going to be a very long discussion over the next year."

The White House plan also underscores the Obama administration's long-standing emphasis on clean energy. Funding for the Department of Energy's (DOE's) clean-energy technology and energy-efficiency programmes would rise by more than $800 million, to $2.7 billion, with sharp increases for clean-vehicle and building technologies and advanced manufacturing.

The DOE's high-risk, high-impact research agency, the Advanced Research Projects Agency—Energy,

would receive $325 million, a $45-million increase from 2015. The department's budget also includes $560 million for research related to fossil fuels, including work on carbon capture and sequestration, and $908 million for nuclear-energy research.

Overall funding in the DOE Office of Science would increase by more than 5%, to $5.3 billion, with most of the increase in basic energy sciences and advanced computing. The budget would slash the fusion-energy programme and prevent payments to the international fusion-reactor project ITER, which is behind schedule and over-budget, unless management reforms are undertaken.

Space and Earth sciences

NASA's budget would increase by $500 million to $18.5 billion, with its science funding holding steady at around $5.3 billion. The agency's Earth-science programme would get a major boost, increasing by roughly 10% to $1.9 billion. Some of that money would be used to begin planning Landsat-9, the next probe in a series of satellites that has monitored land use and land-cover change since 1972. NASA would also assume responsibility for climate satellites currently overseen by the National Oceanic and Atmospheric Administration (NOAA).

The White House plan seeks $1.4 billion for planetary science, a drop of roughly 5%. That pot includes funds to capture and study a small asteroid by moving it into the Moon's orbit. NASA is also requesting $30 million to begin formal planning of a mission to Jupiter's moon Europa. The mission — to a body whose icy crust covers a watery ocean that could, perhaps, support life — has historically won only lukewarm support from the White House but received much stronger backing from Congress, which set aside $100 million for the plan in 2015.

"The increase in the Earth-science budget, coupled with the decrease in the planetary-science budget, is just setting the administration up for a long debate with Congress," says Marcia Smith, a space-programme analyst and founder of SpacePolicyOnline.com.

The NASA request would boost funding for the Stratospheric Observatory for Infrared Astronomy from $70 million in 2015 to $85 million in 2016. The White House had sought to cut funding for the programme entirely last year, but supporters in Congress kept the specially outfitted Boeing 747 flying.

NASA is seeking to end support for two long-running programmes in 2016: the Mars Opportunity rover, which has operated on the red planet since 2004, and the Lunar Reconnaissance Orbiter, which has mapped the Moon's surface since 2009. The agency's chief financial officer, David Radzanowski, says that NASA will decide in the summer whether to reconsider and continue operating the ageing Opportunity, but there is no guarantee that it will set aside money to do so.

NOAA would receive just under $6 billion, up from $5.4 billion in 2015. More than one-third of the budget, about $2.4 billion, would go to the agency's weather and climate probes — including $380 million to develop a mission to avert a potential data gap in the polar-orbiting satellite programme.
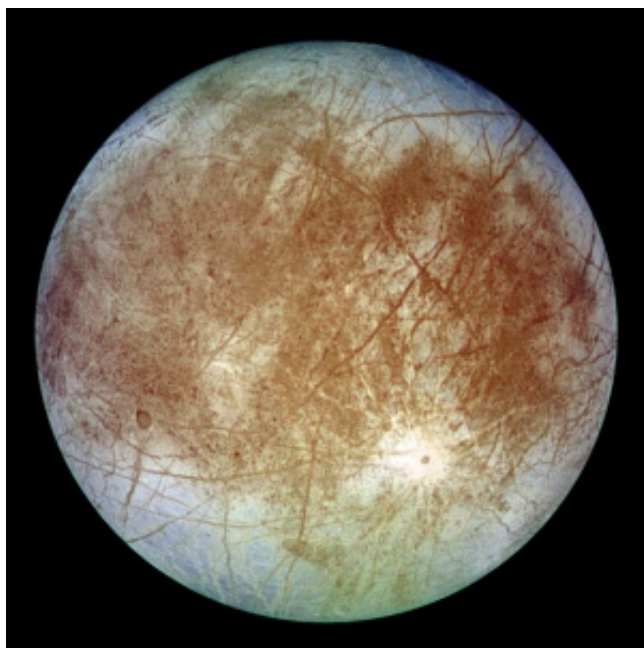
NASA/JPL/DLR

Jupiter's moon Europa, target of a NASA mission.

Obama's budget also revives his failed 2012 proposal to move NOAA from the business-focused Department of Commerce to the Department of the Interior. The interior department has much in common with NOAA: it conducts environmental and climate research, oversees some fisheries and regulates ocean oil and gas drilling. But merging NOAA into the department is likely to remain a tough sell — even within Obama's administration. "We have not been advocating for movement of NOAA," interior secretary Sally Jewell told reporters on 2 February.

The US Geological Survey would see its budget increase by 14% over 2015, to $1.2 billion, with an extra $37.8 million for tools to support land management and a $32-million boost for work on climate resilience and adaptation.

Biomedical sciences

For the first time in years, the US National Institutes of Health (NIH) was slated for a sizeable budget increase — to $31.3 billion, $1 billion more than it received in 2015.

"This is exciting," says Stefano Bertuzzi, executive director of the American Society for Cell Biology in Bethesda, Maryland. "In the scheme of everything, it brings us back to where we were before sequestration."

The White House has been particularly active in creating biomedical research programmes in recent months, requesting $215 million for a new Precision Medicine Initiative that would integrate health and genetic data from at least one million volunteers into a huge database, informing efforts to precisely tailor treatments to individuals (see *Nature* http://doi.org/zvh; 2015). That programme, to the relief of many researchers concerned about funding, will not be paid for out of current research programmes at agencies such as the NIH.

The budget proposal also includes funding to implement a September executive order that charged an interagency working group with developing a national strategy to combat antibiotic-resistant bacteria. The 2016 budget request nearly doubles US government spending on antibiotic-resistance programmes to more than $1.2 billion, with $461 million of that going to the NIH for projects such as developing new antimicrobial agents, and basic research characterizing how resistance evolves. The Biomedical Advanced Research and Development Authority would receive $192 million to develop and

Related stories

- Obama to seek $215 million for precision-medicine plan
- Survey finds US public still supports science
- US lawmakers seek to revamp biomedical

stockpile those new drugs, and $280 million would go towards activities
such as surveillance of resistant strains by the Centers for Disease
Control and Prevention (CDC) in Atlanta, Georgia.

research

More related stories

Current NIH programmes would also benefit under the plan. The agency's budget for its share of the
Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative would more than
double, from $64 million in fiscal year 2015 to $135 million this year.

The NIH budget request is still haunted by the ghost of the longitudinal National Children's Study, which
had received a total of $1.2 billion in funds before the NIH cancelled it in December 2014. Yet the 2016
budget requests another $165 million for that specific study, or for related research to characterize
environmental influences on children's health. Because the study's cancellation was so recent, there
was not time to remove it from the budget, says NIH director Francis Collins. The agency is currently
mulling what to do with the $165 million that was appropriated for the children's study in 2015.

Although the Ebola epidemic in West Africa is waning, the CDC would get $294 million more than in
2015 to study emerging zoonotic infectious diseases, giving a total of $699 million for these activities.
The agency would also receive a $12-million increase for its Global Health Security Agenda, which
would help developing countries to improve surveillance and detect future outbreaks before they get out
of control.

The president's budget also proposed to combine the food-safety offices at the Food and Drug
Administration and the US Department of Agriculture (USDA) into a single agency within the
Department of Health and Human Services (HHS). Ellen Murray, assistant secretary for financial
resources at the HHS, says that the department has worked out few details about the proposed agency.

Agricultural research was another big winner, with an 18% rise over 2015. A significant slice of that
increase is a $450-million request for the USDA's Agriculture and Food Research Initiative. That money
would pay for competitive research grants in areas such as water quality, bioenergy, food safety and
sustainable agriculture.

Basic research

The budget for the National Science Foundation (NSF) would grow to $7.7 billion, roughly a 5%
increase from 2015 and a "very sustainable number for maintaining a healthy research environment",
says Meghan McCabe, a legislative-affairs analyst at the Federation of American Societies for
Experimental Biology in Bethesda, Maryland.

Of note were several new and continuing cross-disciplinary initiatives. The NSF is seeking $144 million
to fund its part of the multi-agency BRAIN Initiative, a 35% increase over what was implemented in
2015. Three other prioritized initiatives would stimulate cross-disciplinary research in food, water and
energy systems; risk and resilience planning; and increased diversity among students in science,
technology, engineering and mathematics.

The amount allocated to new facilities remained flat, and this year, as last year, the NSF did not request funds for new building projects. The agency will continue to fund three ongoing construction projects: the Daniel K. Inouye Solar Telescope, the Large Synoptic Survey Telescope and the National Ecological Observatory Network (NEON). However, the NSF is planning to explore new projects for the future, because construction of NEON will end in fiscal year 2016. The need for research and planning for those future projects is reflected in the NSF facilities budget, which, although declining slightly from $201 million in 2015 to $200 million in 2016, puts 50% more towards concept development and planning.

Another major presidential priority is innovation, particularly in defence, nanotechnology and manufacturing. The Defense Advanced Research Projects Agency, which supports some of the most audacious government-funded research, would increase by $101 million to $3 billion. A total of $1.5 billion would fund work across the government through the National Nanotechnology Initiative. And the president will seek to expand the national network of manufacturing-innovation institutes, something he has pushed for in previous budgets.

Related stories and links

From nature.com

- **Obama to seek $215 million for precision-medicine plan**
  30 January 2015
- **Survey finds US public still supports science**
  29 January 2015
- **US lawmakers seek to revamp biomedical research**
  28 January 2015
- **Obama acts alone on climate**
  27 January 2015
- **Obama urges climate action in State of the Union speech**
  21 January 2015
- **US midterm elections offer little hope for science**
  21 October 2014
- **Obama's budget request falls flat**
  11 March 2014

From elsewhere

- **White House budget request**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments

There are currently no comments.

See other News & Comment articles from *Nature*

Nature   ISSN 0028-0836   EISSN 1476-4687

The South Pole Telescope (left) and the BICEP2 instrument (right) in Antarctica are rivals in the bid to detect signatures from the Universe's first moments.

COSMOLOGY

Gravitational–wave hunt enters next phase

A landmark result has turned to dust, but the search for primordial cosmic ripples continues.

BY ELIZABETH GIBNEY

When the physicists who run the South Pole-based radio telescope BICEP2 found what seemed to be a signal from the dawn of time, it felt like a slam dunk. They were hunting a pattern in cosmic radio waves that would confirm that the Universe had expanded rapidly in the first fraction of a second of its existence — an extremely faint signal if it exists at all. "We all thought this was the kind of problem that you'd have to struggle with for a long time," says Colin Bischoff, a cosmologist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and member of the BICEP2 team. For once, it seemed, nature was being kind.

But it was not to be. Soon after the announcement was made, to great fanfare, in March last year, questions started to emerge. The BICEP2 analysis hinged on a twisted pattern detected in the polarization of the cosmic microwave background (CMB), light left over from the Big Bang. The team attributed the twists, known as B modes, to gravitational waves — ripples in space-time generated during the earliest moments of the Universe. Cosmologists think, but have never proved, that during this period the cosmos underwent

a brief but tumultuous episode of expansion known as inflation. In September, however, results from the European Space Agency's Planck space telescope suggested that dust, which can also emit polarized light, could instead account for the BICEP2 signal.

Last week, those fears were confirmed with the long-awaited announcement of results from a joint analysis carried out by BICEP2 and Planck (see go.nature.com/muyr3z). On 30 January, researchers from both teams described how they had overlaid data recorded by BICEP2 with data from the same patch of sky recorded by Planck at a higher frequency, where the signal is almost entirely attributable to dust. Subtracting the portion of the signal known to be a result of dust left only a tiny excess, with a statistical confidence far below the level needed for a significant finding. There is now no reason to believe that BICEP2 saw anything but dust.

Still, it may yet be possible to find evidence of gravitational waves in the CMB by exploring the existing excess in more detail, and by scanning the sky at different frequencies. The

"Looking back, we probably could have been more cautious."

team has learned lessons from its and Planck's findings, which it is applying in new detectors.

On a bright day in December 2014, as the BICEP2 team was finalizing the joint paper with Planck, Bischoff was in a huge but near-empty hangar at Harvard, finishing upgrades to one of the five telescopes that together make up what is known as the Keck Array, also at the South Pole. Results from the array, which first started taking data in 2011, were rolled into the latest joint analysis, but will also be crucial to the next phase of the gravitational-wave hunt.

Each of Keck's five telescopes is individually as sensitive as the entire BICEP2 telescope, allowing the array to measure the faint signals — detected as tiny temperature differences — with even greater precision. A new role for Keck will be to look for dust: two of its telescopes have been tuned to detect polarized light at a higher frequency than BICEP2, at which dust is expected to polarize the CMB more strongly. A measurement of dust-based polarization more precise than Planck's could give statistical weight to the small excess signal that remains in the BICEP2 data — if that signal is, after all, from primordial gravitational waves.

The array will be assisted by BICEP2's replacement, BICEP3, which packs the same

sensitivity as the Keck Array into a single unit. It has already started collecting data and will be fully operational by the end of February, when the Antarctic summer finishes. Like BICEP2, BICEP3 will look for B modes caused by gravitational waves, but with greater precision and sensitivity, allowing it to detect ever fainter imprints in the CMB. And it will search the sky at 95 gigahertz (GHz), a lower frequency than its predecessor: the joint analysis with Planck, which scans the sky at a variety of frequencies up to 857 GHz, suggested that dust should have less of an effect at 95 GHz, making it more of a "sweet spot" for seeing a primordial signal, says Bischoff.

Antarctica is the perfect place to look for tiny twists in cosmic light. Not only is the continent blessed with relatively clear skies, but its dry climate means that there is much less water in the air to absorb microwaves. Bischoff's colleagues are now working feverishly through the remaining Antarctic summer to get the detectors ready. Even the isolation can be an advantage, says Bischoff: "It's a good place to get work done, and it's pretty beautiful."

But BICEP3 and the Keck Array might be beaten by a rival. They point at the same region of sky as BICEP2, which turns out to be more polluted with dust than once thought. More luck might be had by the South Pole Telescope (SPT), which is less sensitive but scans the sky more widely and at a higher resolution, or the POLARBEAR telescope, installed in 2012 at the James Ax Observatory in Chile, says Peter Coles, a cosmologist at the University of Sussex, UK. "I wouldn't like to pick the likely winner," he adds. The SPT and BICEP teams are also working on a joint analysis: if the primordial signal is very weak, it will be harder to differentiate from another source of B modes known as gravitational lensing, which the SPT is optimized to study.

Anthony Challinor, a cosmologist at the University of Cambridge, UK, is upbeat about the BICEP team's chances. The researchers' growing experience in untangling the CMB and the results obtained at varying frequencies puts them in good stead. "This is a very competitive field and the competition is catching up, but the BICEP team is still ahead of the game."

How did the drama of the discovery affect the team? Acknowledging "some ups and downs" in 2014, Bischoff shrugs his shoulders. "Looking back, we probably could have been more cautious," he says. "But even with a low-key announcement there still would have been a large reaction one way or another." Despite the attention, he says, "I feel like mostly we've kept pretty steady". ∎

Additional reporting by **Ron Cowen**.

Silicene makes its transistor debut

Creation of electronic device using atom–thin silicon sheets could boost work on other flat materials.

BY MARK PEPLOW

Seven years ago, silicene was little more than a theorist's dream. Buoyed by a frenzy of interest in graphene — the famous material composed of a honeycomb of carbon just one atom thick — researchers speculated that silicon atoms might form similar sheets. And if they could be used to build electronic devices, these slivers of silicene could enable the semiconductor industry to achieve the ultimate in miniaturization.

This week, researchers took a significant step towards realizing that dream, by unveiling details of the first silicene transistor[1].

Although the device's performance is modest, and its lifetime measured in mere minutes, this proof of concept has already been causing a stir at conferences, says Deji Akinwande, a nanomaterials researcher at the University of Texas at Austin who helped to make the transistor. Guy Le Lay, a materials scientist at Aix-Marseille University in France, agrees.

"Nobody could have expected that in such a short time, something that didn't exist could make a transistor," he says.

Le Lay was one of the first scientists to create silicene in the lab[2], in 2012 (see 'The rise of silicene'). The debut coincided with a growing sense that graphene was unsuitable for making transistors. Graphene may be the world's most conductive substance, but it is missing a crucial characteristic. Unlike the semiconductors used in computer chips, it lacks a band gap — the energy hurdle that electrons must vault before they can carry a current. Band gaps enable semiconductor devices to switch on and off and to perform 'logic' operations on bits.

"For logic applications, graphene is hopeless," says Le Lay. By contrast, silicene can boast a band gap, because some of its atoms buckle upwards to form corrugated ridges, which puts some of its electrons in slightly different energy states. What is more, makers of electronic chips have been wary of ditching decades of silicon-manufacturing experience in favour of carbon, says Lok Lew Yan Voon, a theoretical physicist at the Citadel, the Military College of South Carolina in Charleston, who first named silicene and modelled its properties back in 2007 (ref. 3).

But handling silicene in the lab has been a

THE RISE OF SILICENE

Its carbon-based cousin graphene gets more attention, but silicene is catching up.
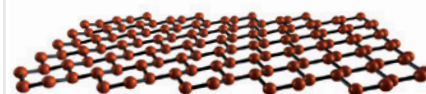
1994 First calculation of the structure of two-dimensional crystals of silicon (pictured) and of germanium.

2004 Andre Geim and Konstantin Novoselov report isolation of graphene.

2007 The name 'silicene' is coined.

2009 Fabrication of silicene nanoribbons; flurry of theoretical papers on silicene and germanene begins.

2010 Geim and Novoselov bag Nobel Prize in Physics for their experiments on graphene.

2012 Six independent reports of silicene sheets (formed on silver).

2015 First demonstration of silicene transistor.

huge challenge. The material cannot be peeled from a solid block using sticky tape, as graphene can from bulk graphite. Instead, researchers produce it by letting a hot vapour of silicon atoms condense onto a crystalline block of silver in a vacuum chamber, a much trickier process. And unlike robust graphene, naked silicene is extremely unstable in air, making it difficult to transfer the gossamer sheet to more useful substrates — such as the guts of a transistor. As recently as last year, some researchers were still questioning whether silicene even existed.

So Akinwande joined forces with Alessandro Molle at the Institute for Microelectronics and Microsystems in Agrate Brianza, Italy, to offer silicene some protection. They formed a silicene sheet on a thin layer of silver, and added a 5-nanometre-thick layer of alumina ▸

▶ on top. Then they peeled this silicene sandwich off its mica base, flipped it silver-side-up, and laid it on an oxidized-silicon substrate. Finally, they gently etched away some of the silver to leave two islands of metal as electrodes, with a strip of exposed silicene between them.

"It's a very clever trick," says Le Lay, who is planning to try the process with germanene, a capricious, similarly structured 'two-dimensional' material made from germanium that his team created last year[4].

Clever it may be, but the transistor will not be making an appearance in mobile phones any time soon: the exposed silicene degrades in about two minutes. Still, that is long enough to measure its properties. Although its electrons are sluggish in comparison to graphene's, the device does indeed have a small band gap.

Laying an extra coating on top of the silicene transistor could also extend its life. Akinwande has used Teflon to help flakes of phosphorene — another air-sensitive, two-dimensional material, made of phosphorus — to survive for months[5]. Other researchers have shown that using multiple layers of silicene could allow the sacrificial top layers to protect those beneath for 24 hours[6]. Crucially, the technique used to make the silicene transistor could now help to test all of these ideas, and more, with various air-sensitive materials. "It's definitely a game-changer," says Lew Yan Voon. "This is the paper we've been waiting for in the field."

"It's definitely a game-changer. This is the paper we've been waiting for."

Not everyone is so enthusiastic about silicene's prospects. "There's a lot of talk about silicene, germanene and phosphorene," says Jari Kinaret of Chalmers University of Technology in Gothenburg, Sweden, who is the director of the European Union's Graphene Flagship, a €1-billion (US$1.1-billion) research project to study two-dimensional materials and develop applications for them, "but the difficulties with them are still quite substantial."

Le Lay, however, is convinced that researchers will flock to silicene. "Now that a device has been made," he says, "other scientists will see it is not a dream material, it is a practical thing." ∎

1. Tao, L. *et al.* *Nature Nanotechnol.* http://dx.doi.org/10.1038/NNANO.2014.325 (2015).
2. Vogt, P. *et al.* *Phys. Rev. Lett.* **108,** 155501 (2012).
3. Guzmán-Verri, G. G. & Lew Yan Voon, L. C. *Phys. Rev. B* **76,** 075131 (2007).
4. Dávila, M. E., Xian, L., Cahangirov, S., Rubio, A. & Le Lay, G. *New J. Phys.* **16,** 095002 (2014).
5. Kim, J.-S. *et al.* Preprint at http://arxiv.org/abs/1412.0355 (2014).
6. De Padova, P. *et al.* *2D Mater.* **1,** 021003 (2014).

The King Abdullah University of Science and Technology is a stark exception to strict Saudi society.

SAUDI ARABIA

Science oasis under pressure

But scientists at Saudi Arabia's leading university argue for a quiet approach to modernizing the nation.

BY DECLAN BUTLER

Following the high-profile flogging of Saudi Arabian activist Raif Badawi last month, the King Abdullah University of Science and Technology (KAUST) — a multi-cultural, world-class university in what seems an unlikely setting — is in the spotlight.

Badawi received 50 lashes, the first in a sentence that stipulates a total of 1,000 lashes plus 10 years in prison, as punishment for a website that he created for social and political discussion. As well as prompting an international outcry, the case has put KAUST's leaders under pressure to speak out about the lack of freedom of expression in Saudi Arabia, where KAUST is based. Researchers at the university, however, argue that they can have a bigger impact on Saudi society — and perhaps on the Arab and Muslim world broadly — by quietly continuing in their efforts to create a world-class centre for research and critical thinking.

"KAUST is built on values that I espouse as a scientist, and the impact of KAUST will be felt over time, in major part through the influence of its graduates," says Mark Tester, an Australian who is associate director of KAUST's Center for Desert Agriculture.

A graduate university, KAUST was founded in 2009 by the late King Abdullah, with the goal of establishing a culture of science and enlightenment in Saudi Arabia and beyond.

A stark exception to strict Saudi society, its campus in Thuwal, 90 kilometres north of Jeddah, imposes no discrimination on the basis of sex, religion or ethnicity. Unlike in the rest of the country, women and men mingle, and women can also drive. The freedoms on the campus were a condition of the prominent Western scientists who backed KAUST's development.

On 18 January, nine days after Badawi received the first lashes, 18 Nobel prizewinners from around the world wrote to Jean-Lou Chameau, the president of KAUST, calling on "influential voices in KAUST" to speak up for the freedom to dissent. The letter warns that KAUST's international ties could be at risk if the restrictions on freedom of thought and expression in Saudi Arabia continue.

One researcher familiar with KAUST, who requested anonymity because of the sensitivity of the issues, says that if KAUST researchers were to speak out or be politically active, it would have little effect on the regime and would risk providing ammunition for the institution's critics in Saudi Arabia. KAUST is controversial there, the researcher says,

SUSAN BAAGHIL/REUTERS

and the state and clerics have sought to keep the university and its scientists at a distance from domestic political or social issues. "It is always under a microscope from conservative elements," says the researcher.

Scientists can do more for Saudi Arabia by working at KAUST than by criticizing it from outside, says Tester. "We are making a real contribution to the country through education, and through research advances."

KAUST has attracted leading scientists from around the world to join its faculty of around 130, and has set up science centres to study regionally important issues such as desert agriculture, Red Sea research, desalination and solar energy. The campus now hosts 840 students from 69 countries, including 246 from Saudi Arabia and 302 women.

"My philosophy is that I don't think I'm compromising, but modestly contributing to opening up things," says another foreign researcher who requested anonymity.

Indeed, much of the international support that was crucial for KAUST's development came with the understanding that Saudi Arabia would improve freedoms beyond the campus site. Yet, as the case of Badawi highlights, if anything, the kingdom seems to have stepped up its repression of freedoms since KAUST was founded.

There was a "spirit of hope" when KAUST opened, says letter co-signatory John Polanyi, who won the 1986 Nobel Prize in Chemistry. But patience with the Saudi Arabian regime is now "wearing thin". "I think the scholarly community has been slow to become aware that KAUST cannot be an island of freedom," he says.

Tester argues that KAUST is educating a new generation of Saudi students, who will eventually help to transform the kingdom more generally. "KAUST's existence is evidence of the kingdom's desire to develop," he says. "It will take time, and I ask that people give us time."

KAUST is not the only academic force for change in Saudi Arabia. A multibillion-dollar scholarship programme launched by King Abdullah in 2005, and set to continue until 2020, funds hundreds of thousands of Saudi undergraduates and postgraduates to study abroad. Scientists familiar with Saudi Arabia say that they suspect Abdullah's plan was to produce a delayed benefit: after being exposed to alternative ideas and cultures, returning students would moderate Saudi society and ease the grip of conservative clerics. Education in the

"We are making a real contribution to the country through education, and through research advances."

kingdom is heavy on religion. "It's more Koran than periodic table," says one researcher.

Although large numbers of Saudi men have long had Western educations, one big difference that the programme provides is that it is open to women. "That is what will be transformative," says another foreign scientist who has worked closely with KAUST. "But it's not going to happen overnight."

Still, scientists inside and outside KAUST agree that the establishment of a knowledge-based culture and economy will require reforms by the Saudi leadership too. "The whole idea behind KAUST was that King Abdullah wanted Saudi Arabia brought back into the mainstream of science," says one anonymous scientist. But modern science requires free thinking and creativity, and cannot flourish in a repressive culture, adds the researcher. "If Saudi Arabia is to take its place on the modern science and technology scene it really has to pay attention to its human rights." ∎

CORRECTION
The News story 'Rave drug tested against depression' (*Nature* **517,** 130–131; 2015) stated that ketamine acts by blocking the signalling molecule NMDA. The drug actually acts on the NMDA receptor.

NO DRIVERS REQUIRED

Automation is one of the hottest topics in transportation research and could yield completely driverless cars in less than a decade.

BY M. MITCHELL WALDROP

PHOTO: JOHN CHAPPLE/SPLASH NEWS/CORBIS. SOURCES: WHO; NATL HIGHWAY TRAFFIC SAFETY ADMINISTRATION; CENTER FOR INTERNET AND SOCIETY

PASCAL GOETGHELUCK/SPL

This summer, people will cruise through the streets of Greenwich, UK, in electric shuttles with no one's hands on the steering wheel — or any steering wheel at all. The £8-million (US$12-million) project, part of a larger study of driverless cars funded by the UK government, is just one of many efforts that seek to revolutionize transportation. Spurred in part by a desire to end the carnage from road accidents — about 90% of which are caused by driver error — the race is on to transfer control from people to computers that never doze at the wheel, get distracted by text messages or down too many pints at the pub.

Almost every major car maker is working on some form of automation, as are many electronics companies. But looming over everyone is the Internet giant Google: the company has been widely acknowledged as the world leader in driverless-car research since October 2010, when it announced that it had entered the field a year earlier — and that its driverless test vehicles had already logged more than 200,000 kilometres on roads near its headquarters in Mountain View, California, and elsewhere in the state. The public's enthusiastic response to that revelation galvanized car makers and government research-funding agencies around the world to accelerate their efforts in this arena.

"I've never seen anything move so quickly from concept into products," says Richard Bishop, an automotive consultant who headed a US Department of Transportation research programme on automated motorways in the 1990s. Although many technical challenges remain, developers say they can see clear paths for solving most or all of them.

At Google, for example, most of the driverless-car work so far has been carried out using standard passenger vehicles fitted with Global Positioning System receivers and mapping technology, along with radar to detect obstacles, a laser ranging system to scan the surroundings in three dimensions, and video cameras to identify objects such as traffic lights, construction signs, pedestrians and other vehicles (see 'A world of driverless cars'). The on-board computer — with processing power equivalent to several desktop units — integrates all the information and decides how the car should behave in any given situation. To lessen the load on the driving algorithms, Google equips the car with ultra-detailed maps that tell it exactly what to expect, down to the height of every curb.

Sceptics point out that this mapping requirement restricts the car to places that have already been surveyed to that level of precision, such as Mountain View. But it would be relatively easy to expand those maps as part of the company's ongoing efforts to photograph the world's roadways for Google Maps, says Sebastian Thrun, an engineer who founded the Google car project and ran it until 2013.

A tougher problem, says Thrun, is teaching the car how to respond to what he calls "the long tail of unlikely events". Early on, he says, the Google team developed algorithms for handling frequent, obvious challenges such as intersections or rain-slicked roadways. But as the cars drove for thousands of kilometres, they recorded oddball events such as a plastic bag blowing across the motorway or a couch sitting in the middle of the road. "There were many more of those than we believed in the beginning," says Thrun. The only way to handle such rare events has been to record them as they arise, devise responses with the help of high-powered machine-learning algorithms — and then test those solutions with simulations and yet more driving.

"If we do it long enough," says Thrun, "the hope is that the software will be as safe as a human driver" — and eventually much safer. How long that will take remains an open question. Google has publicly estimated about five years — but the company is currently not granting interviews about its project.

Additional safety could come from equipping cars and trucks with Wi-Fi-like vehicle-to-vehicle (V2V) radios, which would allow them to warn each other of dangerous situations such as a car running through a red light. That would give both driverless and human-operated vehicles time to steer clear.

Although V2V technology will probably be incorporated into driverless vehicles, it has been developed largely through separate efforts. The concept has already been road-tested as part of the European Union's Safe Road Trains for the Environment project, in which lines of cars followed bumper to bumper behind a truck, like ducklings tailing their mother. These road trains, or platoons, avoid catastrophic pile-ups because the V2V signals cues every car to hit the brakes at the same instant as the truck. And because of aerodynamics, the road trains saved at least 10% in fuel consumption.

Such experiments have piqued the interest of the car maker General Motors, which last September announced it will support V2V technology in future models. At first, such cars will have few other vehicles to talk to. But the US National Highway Traffic Safety Administration is looking to issue regulations requiring V2V radios in all new US cars later this decade.

"The vehicles would just be broadcasting a basic safety message: position, speed, direction of travel," says Josh Switkes, chief executive of Peloton Technology, a start-up firm in Menlo Park, California, that is seeking to commercialize V2V technology for heavy trucks. But that will be enough to eliminate many accidents, he says. Then there are the energy efficiencies — not just from platooning, he says, but from cars working together to minimize stop-and-go traffic, where fuel efficiency is dismal. If the cars can also ask smart traffic lights to adjust themselves to prevailing traffic density — a practice sometime known as Vehicle to Infrastructure (V2I) — then the system might be able to minimize the need to stop at all.

Ultimately, the timescale for deploying these technologies will depend on the answers to much broader questions. How much will they cost? Who will own them — individuals, or service companies that provide transportation on demand? Who will face legal liability when a driverless car gets into an accident? And will people accept and trust them?

Such questions can only be answered through experience. And given the pace of innovation today, experience is accumulating fast. ∎

M. Mitchell Waldrop *is a features editor for* Nature *in Washington DC.*

DITCH THE DRIVER

1.24 million
traffic fatalities every year worldwide

90%
of all accidents are due to driver error

4 US states
and the District of Columbia have passed laws to allow driverless cars on their roads

A world of driverless cars

Fully autonomous vehicles are developing faster than anyone would have thought a few years ago, with many experts predicting that they will become widely available in the next 5–10 years. Many questions remain, but it is already possible to imagine how this new world of driverless cars will work.

PERCEPTION
Vehicles use **radar** to detect obstacles, a **laser ranging system** to map the surroundings in three dimensions, and **video cameras** to identify objects such as traffic lights, construction signs, pedestrians and other vehicles.

DECISION AND ACTION
To make the appropriate responses to rare events — such as a ball bouncing in from a playground, or a plastic bag blowing down the roadway — the cars rely on **algorithms** refined through millions of kilometres of test drives.

ADAPTIVE TRAFFIC FLOW
Smart infrastructure integrates V2V signals from the moving cars to optimize speed limits, traffic-light timing and the number of lanes in each direction on the basis of the actual traffic load. The result is a smoother flow, shorter travel time and less energy wasted at traffic lights or in traffic jams.

COMMUNICATION
Vehicle-to-vehicle (V2V) radios send signals between cars, trucks and infrastructure items such as traffic lights.

ROUTE PLANNING
An **on-board computer** uses sensor data to plot a route that gets the car where it needs to go, while avoiding people, potholes and other vehicles.

LOCATION
Mapping software uses **Global Positioning System** data to tell the car where it is in relation to roads, traffic signals, and other landmarks.

2020s
The decade when driverless cars are predicted to become widespread.

ILLUSTRATION BY DON FOLEY; TEXT BY M. MITCHELL WALDROP; DESIGN BY KELLY KRAUSE

10%
Fuel savings for cars that travel in formation.

ROAD TRAINS
Vehicles can take advantage of aerodynamics and save fuel by following one another almost bumper to bumper. They are protected from catastrophic pile-ups by their **V2V radios,** which allow all the cars in line to hit their brakes at the same time.

CITIES TRANSFORMED

MASS TRANSPORT People increasingly give up owning cars in favour of calling companies to pick them up wherever they are and drop them off wherever they need to go — a driverless version of a ride-sharing service.

LAND USE Urban centres begin to undo the many accommodations they have made for personal vehicles — starting with the vast quantities of real estate devoted to parking, which could be adapted to more productive uses.

800 million
One estimate of the number of US parking spaces. Many could be used for other purposes if people ride-share more.

THE BRAIN,
INTERRUPTED

Babies are increasingly surviving premature birth — but researchers are only beginning to understand the lasting consequences for their mental development.

BY ALISON ABBOTT

Fabienne never found out why she went into labour three months too early. But on a quiet afternoon in June 2007, she was hit by accelerating contractions and was rushed to the nearest hospital in rural Switzerland, near Lausanne. When her son, Hugo, was born at 26 weeks of gestation rather than the typical 40, he weighed just 950 grams and was immediately placed in intensive care. Three days later, doctors told Fabienne that ultrasound pictures of Hugo's brain indicated that he had had a severe haemorrhage from his immature blood vessels. "I just exploded into tears," she says.

Both she and her husband understood that the prognosis for Hugo was grim: he had a very high risk of cerebral palsy, a neurological condition that can lead to a life of severe disability.

The couple agreed that they did not want to subject their child to that. "We immediately told the doctors that we did not want fierce medical intervention to keep him alive — and saw the relief on the doctors' faces," recalls Fabienne, who requested that her surname not be used. That night was the most tortured of her life.

The next day, however, before any change had been made to Hugo's treatment, his doctors proposed a new option to confirm the diagnosis: a brain scan using magnetic resonance imaging (MRI). This technique, which had been newly adapted for premature babies, would allow the doctors to predict the risk of cerebral palsy more accurately than with ultrasound alone, which has a high false-positive rate. Hugo's MRI scan showed that the damage caused by the brain haemorrhage was limited,

and his risk of severe cerebral palsy was likely to be relatively low. So just 24 hours after their decision to let his life end, Hugo's parents did an about-turn. They agreed that the doctors should try to save him.

Thanks to medical advances since the 1970s, premature infants — those born before 37 weeks of gestation — are increasingly able to survive. Some hospitals now try to save babies born as early as 22 weeks. But those developments are forcing doctors and parents to grapple with difficult decisions, because the chances of severe disability increase with the extent of prematurity. Cerebral palsy, for example, affects 1–2% of babies born at term, 9% of those born earlier than 32 weeks and 18% of those born at 26 weeks.

That is just half the story. Neuroscientists are developing an increasingly sophisticated picture of premature infants' brains that could help to inform medical decisions and treatments. From some long-term studies, they are learning that premature children face a higher risk than was previously thought of developing cognitive or behavioural problems — according to some studies, as many as half of them will.

Researchers are starting to ask why this should be, whether it could be avoided and what is the best way to provide educational support for the affected children. "We need to gather a lot more data to understand what the best strategies are," says Petra Hüppi, a neonatologist and developmental paediatrician at the University of Geneva in Switzerland, who is following the brain development of children who were born prematurely.

EARLY BIRTHDAY

Prematurity — also called pre-term birth — is extremely common. According to World Health Organization statistics from 2012, more than one in 10 babies — around 15 million in total — are born prematurely each year. The great majority are born between 32 and 37 weeks of gestation, but 1.6 million are born between 28 and 32 weeks and 780,000 are born 'extremely pre-term', before 28 weeks (see 'Born too soon').

In low-income countries, more than 90% of extremely pre-term babies born alive soon die, which helps to explain why prematurity is now the second biggest cause of death in children under five, after pneumonia. But in richer countries, with sophisticated neonatal intensive-care facilities, more than 90% of these extremely pre-term babies survive, and doctors are continuing to push the age of survival even earlier in development. Doctors in the United States are debating a controversial recommendation to lower the gestational age at which a baby should be considered potentially viable from 24 weeks to 23 weeks. In Japan, babies born at 22 weeks have been considered viable since 1991.

Parents of premature children face agonizing waits as their children fight for their lives.

Hugo's parents endured tense weeks during which their son had a series of operations to fix damaged organs, and to create essential connections between major blood vessels that had not had time to develop before birth. They knew he could still die at any time. "But I felt like we were back on the TGV," says Fabienne, referring to the French high-speed trains. "The train goes fast and it rocks frighteningly — but we were on it again."

But what happens after the immediate danger has passed? Just a few studies have so far followed up the long-term fate of premature babies, because it is time-consuming and

"We were shocked to see just how many children had problems."

expensive to track them with sophisticated cognitive and behavioural tests over many years.

One of the first studies to show the extent of developmental problems was EPIPAGE, which looked at a cohort of all live births between 22 and 32 weeks of gestation from 9 regions of France in 1997, and a reference group of 664 full-term babies[1]. Up to half of the premature babies who survived to five years of age had some sort of neurodevelopmental problem by then, and the impairments in cognitive development grew more pronounced for each extra week of prematurity. On a score of cognitive ability, the team observed impairment in 44% of those born between 24 and 25 weeks of gestation and 26% of those born at 32 weeks, compared with 12% of full-term controls.

"We were shocked to see just how many children had problems," says Hüppi. Moderately premature babies may be at lower risk than extremely premature babies, she notes, but there are many more of them.

The effects seem to continue into adulthood. Developmental psychologist Dieter Wolke led an unusual study of hundreds of children born between 26 and 31 weeks of gestation in Bavaria in the mid-1980s. He assessed them at six years old[2], and again at 26 years[3]. Last year, he reported[3] that most of those who had cognitive problems as children still had them as adults: one-quarter of them had moderate to severe cognitive deficits, and half had mild cognitive deficits. Most of those who experienced problems had short attention spans, and as a group they tended to underachieve academically and career-wise.

Wolke, who is currently at the University of Warwick, UK, observed subtler lifestyle

differences, too. "They are less likely to take risks, smoke, drink or have early sexual relationships," he says.

Scientists are still struggling to understand the physical changes in the brain that underlie all these differences. The brain is made up of grey matter, which comprises densely packed cell bodies, and white matter, the long-reach axons of cells that connect different brain regions. These axons are covered in a protective coating called myelin during development, in a precise sequence that begins in the womb and continues for the first decade or so after birth.

In the premature brain, immature, fragile blood vessels struggle to provide tissue with enough oxygen for normal development. When a vessel ruptures, crucial areas of white matter are destroyed and cerebral palsy can result. But very little is known about what causes the more subtle brain problems that cohort studies of premature infants are revealing.
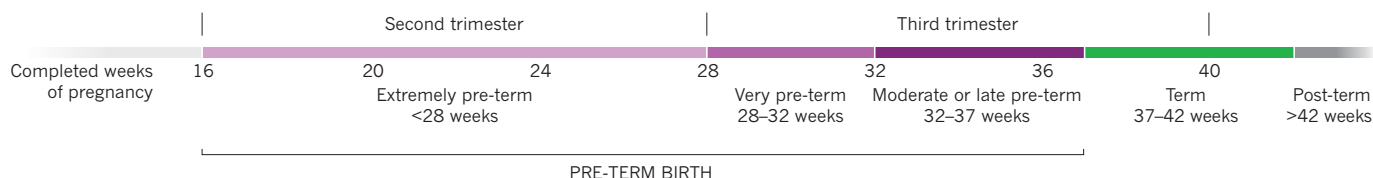
TOO MUCH TOO SOON

Scientists suspect that when the brain is forced to carry out a crucial part of its development while the child is in the outside world instead of a warm, watery womb, it receives inappropriate signals from the environment that affect how its neurons are linked into networks. "The premature brain gets subjected to quite different sensory inputs — like visual stimulation and gravity effects — which it is not supposed to be subject to," says Ghislaine Dehaene-Lambertz of the INSERM-CEA Cognitive Neuroimaging Unit in Paris, who studies language development in infants. "They can be sudden, intense but also unpredictable." Some of these unnatural sensory signals are inevitably provided by the intensive medical procedures that keep premature babies alive.

Pioneering brain-scanning studies support the idea that altered networks play a part in cognitive problems. Hüppi's Swiss collaboration looked at 52 six-year-olds who had been born prematurely, using MRI scans optimized to reveal tracts of neurons connecting brain regions[4]. Compared with children born at term, the premature children's neuronal tracts were organized less efficiently, often taking a more meandering path. These changes in organization were correlated with reduced social and cognitive skills.

In another study, neonatologist Jeffrey Neil, then at St. Louis Children's Hospital in Missouri, and his team used functional MRI to study the premature brain at rest. The low-level, idling activity of a resting brain gives a read-out of its working connections, whose general topology is laid out before birth (see *Nature* **489,** 356–358; 2012). The team showed[5] that in babies born between 23 and 29 weeks of gestation, this 'resting-state connectivity' tends on average to be less complex and active at term-equivalent age than it is in full-term babies at birth. Another study — on the 26-year-old Bavarians — showed[6] that this

BORN TOO SOON

Just over 11% of live births worldwide are pre-term — before 37 weeks of gestation — and premature birth is the second largest cause of child deaths in under-fives. The medical risks increase with the extent of prematurity; neuroscientists now think that some effects on brain development may last into adulthood.

	Second trimester				Third trimester			

Completed weeks of pregnancy	16	20	24	28	32	36	40	
		Extremely pre-term <28 weeks			Very pre-term 28–32 weeks	Moderate or late pre-term 32–37 weeks	Term 37–42 weeks	Post-term >42 weeks

PRE-TERM BIRTH

22 WEEKS
Some hospitals now try to save babies born as early as 22 weeks.

22–32 WEEKS
One long-term study found that up to half of children born in this window have some neurodevelopmental problem at age five.

23–39 WEEKS
Brain-scanning studies point to atypical structural and functional connections in the brains of premature infants.

24 WEEKS
50% chance of survival with neonatal intensive care in most high-income countries.

26 WEEKS
Cerebral palsy affects around 10% of very pre-term babies, but 18% of those born at 26 weeks.

34 WEEKS
50% chance of survival with neonatal intensive care in many low-income countries.

---

reduced complexity of resting-state connectivity stretches into adulthood.

Researchers agree that the most revealing studies would monitor the brains of premature babies and full-term comparison babies from as early as possible after birth, with follow-up scans and assessments throughout life. But such studies are difficult, and not only because it is hard to keep tabs on families who may move house, lose interest or lose touch over the years. Parents are rarely keen for their newborns — whether premature or full-term — to be whisked away into the loud and lonely chamber of a distant MRI machine without a burning medical reason. (In some countries, such as the Netherlands, it is illegal to do so.) And not all obstetricians are comfortable with subjecting delicate premature babies to brain scans at a medically and emotionally fraught time. Fabienne was happy for Hugo to be scanned, but recalls how painfully long it took to get from the paediatric ward to the scanning suite in her hospital. "I felt half-dazed walking alongside Hugo in his incubator through a long underground tunnel to get there," she says. "It looked like the tunnel you are supposed to see when you are dying."

A small vanguard of scientists and clinicians is pushing ahead, and several large long-term studies are under way around the world, collecting neurological, cognitive, behavioural and genetic data from birth, along with brain scans.

In France, EPIPAGE 2 is now running, and has recruited more than 4,200 premature babies from all over the country[7]. In the United Kingdom, a team led by neonatologist David Edwards of King's College London has launched a study that will track children from their time *in utero* until they are two years old, collecting brain scans and blood samples along the way. Some of these children will inevitably be born prematurely, and the plan is to identify molecular signatures that might predict which of those infants are particularly vulnerable, or resistant, to altered neurodevelopment.

Edwards' preliminary studies[8] on premature babies suggest that some genes — including several associated with lipid metabolism, which is crucial for white-matter development — may modify the risk of altered brain development. "Having a particular genetic profile might make certain babies less vulnerable," he says.

BRAIN PROTECTION

With scientists still working to identify the molecular, cellular and network differences in the premature brain, finding treatments seems a fond hope. But Hüppi is attempting to do it. She is conducting a clinical study of erythropoietin, or EPO, a drug that stimulates the production of red blood cells. It is already a standard treatment to aid oxygenation of internal organs — not to mention being a favourite among endurance-sport cheats — and it is also thought to protect and support neurons.

Anecdotal reports had suggested that erythropoietin might help long-term neurodevelopment, and Hüppi's team is assessing this in a prospective, randomized and controlled study in nearly 500 very premature babies born in Switzerland, who are being MRI scanned at term-equivalent age. The first results, published in 2014, showed[9] that treated babies had fewer signs of neurological problems than did children in a control group. But the acid test, says Hüppi, will come when they are assessed at two years old, when neurodevelopment has proceeded further.

Where does all this leave parents, who still have to make decisions about their children's treatment with only limited information about the long-term prognosis? Some, such as Fabienne, can be helped by MRI scans that can detect damage in white and grey matter, and make it possible to predict the risk of severe brain damage more precisely than in the past. Hüppi says that the technology helps doctors to advise parents, "and it is a terrible responsibility if we are wrong". But this does little to identify which children will have milder developmental problems, or what those might be.

Edwards and others think that brain imaging alone can never provide that type of information — but that combining scans with genetic and other molecular and clinical data may eventually lead to much greater precision. Should this become possible, it will throw open a whole new debate about how best to ameliorate any future problems for premature children, through very specific social and educational support — something that neuroscientists and education experts are only beginning to grapple with.

Fabienne, like many parents of premature children, would like to have that information. Hugo, who is now seven, occupies most of her time. He has difficulties with fine movements, and some visual problems; he also needs a lot of extra assistance at school. Fabienne is deeply engaged with educational training programmes, which she hopes will be helpful, although she cannot know for sure. But Hugo is an unadulterated joy to her, and she is endlessly grateful for the MRI scans that were so crucial in the decision to save him. "Neuroscience was able to say that Hugo would be able to have a reasonable quality of life," she says.

And she monitors from a distance the new wave of scientific interest in the brains of premature babies. "Neuroscience is coming up with a lot of good information — I really hope that they will soon translate what they are discovering into concrete actions that parents can usefully undertake." ∎

Alison Abbott *is* Nature's *senior European correspondent.*

1. Larroque, B. *et al. Lancet* **371,** 813–820 (2008).
2. Wolke, D. & Meyer, R. *Dev. Med. Child Neurol.* **41,** 94–109 (1999).
3. Eryigit Madzwamuse, S., Baumann, N., Jaekel, J., Bartmann, P. & Wolke, D. *J. Child Psychol. Psychiatry* http://dx.doi.org/10.1111/jcpp.12358 (2014).
4. Fischi-Gómez, E. *et al. Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhu073 (2014).
5. Smyser, C. D. *et al. Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhu251 (2014).
6. Bäuml, J. G. *et al. Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhu133 (2014).
7. Ancel, P.-Y., Goffinet, F. & EPIPAGE 2 Writing Group *BMC Pediatr.* **14,** 97 (2014).
8. Boardman, J. P. *et al. Pediatrics* **133,** e1655–e1663 (2014).
9. Leuchter, R. H.-V. *et al. J. Am. Med. Assoc.* **312,** 817–824 (2014).

COMMENT

ILLUSTRATION BY PETE ELLIS/DRAWGOOD.COM

PHARMA   RESEARCH
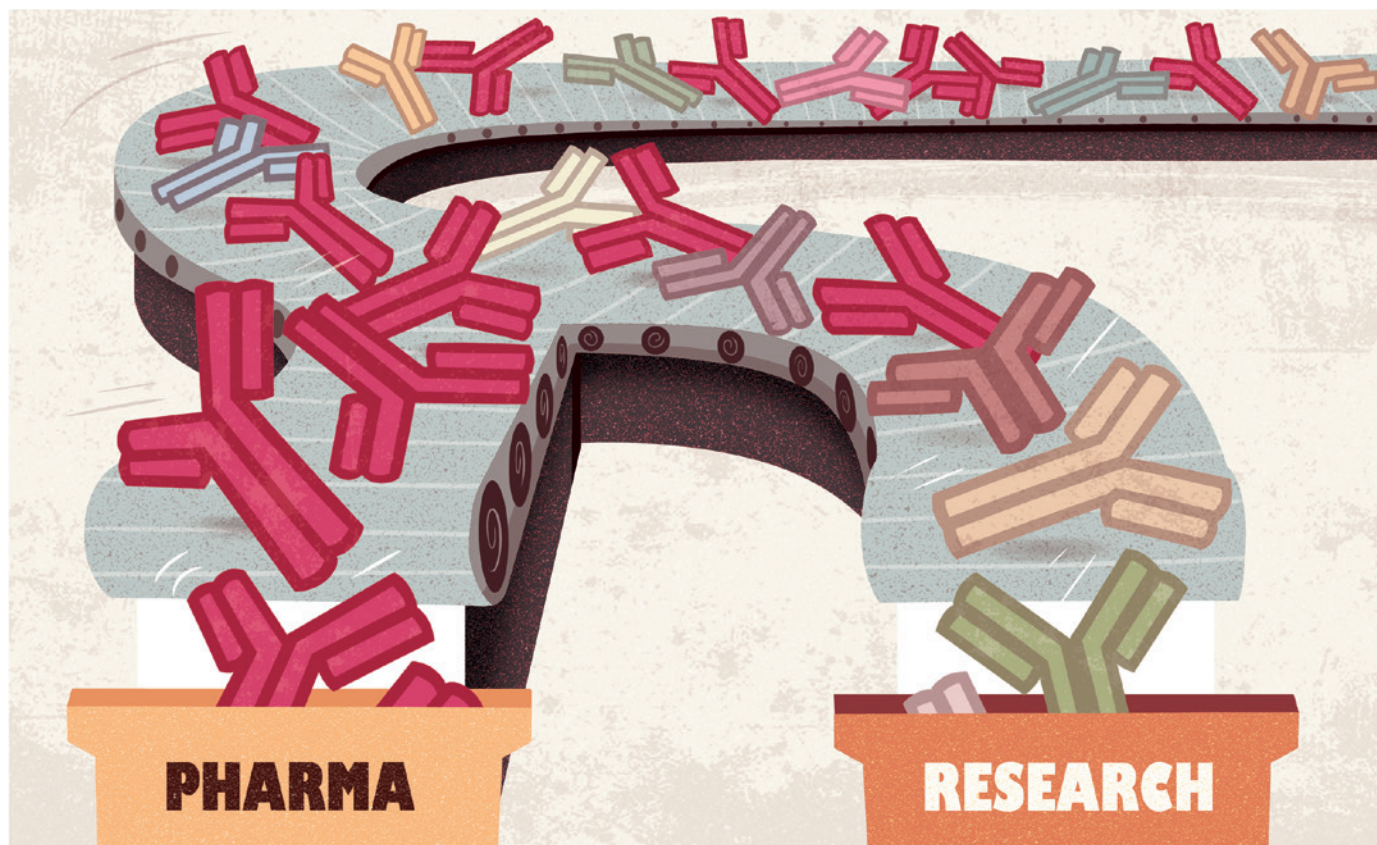
Standardize antibodies used in research

To save millions of dollars and dramatically improve reproducibility, protein-binding reagents must be defined by their sequences and produced as recombinant proteins, say **Andrew Bradbury**, **Andreas Plückthun** and 110 co-signatories.

Central to reproducibility[1] in biomedical research is being able to use reagents that are identical to those described in publications. Alarmingly, there are serious flaws in the reliability of antibodies, the most widely used class of protein-binding reagent[2,3].

In the body, antibodies help to fight pathogens. In the lab, biologists have long used them to track proteins of interest because they bind to specific targets. But in a 2008 study[3], fewer than half of around 6,000 routinely used commercial antibodies recognized only their specified targets, with some manufacturers producing consistently good antibodies, and others consistently poor ones.

This figure may be optimistic[4]. In fact, we believe that poorly characterized and ill-defined antibodies were in large part to blame for a study co-authored by C. Glenn Begley (a co-signatory to this article) being able to replicate the scientific results of only 6 of 53 landmark preclinical studies[5]. Across biomedical research, the resulting waste in materials, time and money is vast — costing an estimated US$350 million annually in the United States alone.

To stem this loss, we call for an international collaboration and funding initiative to define all binding reagents ▶

▶ according to the sequences that encode them. Crucially, researchers should use recombinant antibodies or binding reagents. These are made from reliable cell lines by isolating and incorporating the genes into plasmid DNA and transferring the plasmids into cells or bacteria for culture.

WILDLY VARIABLE

Researchers have used polyclonal antibodies for decades. These are produced by injecting a target (typically a protein) into an animal such as a rabbit and using the resulting serum as a source of antibodies. However, only 0.5–5% of the antibodies in a polyclonal reagent bind to their intended target. And functionality varies from batch to batch, because immunizing an animal — even the same one — never results in exactly the same mix of antibodies, making it hard for researchers to be sure of the specificity of any particular batch of binding reagent obtained in this way.

Four decades ago, the first monoclonal antibodies were made — by fusing a normal antibody-producing B-lymphocyte cell with a cancer cell to produce a 'hybridoma'[6]. The biomedical community believed that the resulting cell lines that produced (ideally) a single antibody species would solve many of the challenges of polyclonals. Unfortunately, monoclonals are far from problem-free.

Hybridoma cell lines can die off, lose their antibody genes, or simply not grow when taken out of frozen storage — meaning that the source of a particular monoclonal antibody may be lost forever. Furthermore, such antibodies may bind to more than one target, either because the antibody is actually a mixture of antibodies with multiple specificities, or simply because it is able to bind to several proteins. Careful characterization is thus required.

Most pharmaceutical and large biotechnology companies have whole departments dedicated to validating and characterizing antibodies. Consequently the reagents used in most clinical trials and especially in medical procedures cleared by the US Food and Drug Administration or the European Medicines Agency are extremely reliable.

Outside clinical trials, reagents are rarely validated to the same degree. What is more, only 44% of publications provide enough information — for instance, on the supplier — for researchers to be able to purchase the same antibody[7]. The quality of the documentation that accompanies batches (such as on functionality in different assays) is enormously variable[2]; even when it is provided, it may not correspond to the batch supplied[8].

TWO STEPS

If all antibodies were defined by their sequences and made recombinantly, researchers worldwide would be able to use the same binding reagents under the same conditions. Immortal production lines of recombinant antibodies — which express no extra antibody chains — can be engineered by incorporating plasmids containing antibody DNA into cell lines.

In practice, improving the quality of protein-binding reagents will require two steps. First, the sequences should be obtained for widely used hybridoma-produced monoclonal antibodies. These antibodies should thence be produced recombinantly. (Polyclonal antibodies should be phased out of research entirely.)

Second, the research community should turn to methods that directly yield recombinant binding reagents that can be sequenced and expressed easily. These include display and two-hybrid methods — in which the best binders are selected from billions of variants — as well as approaches in which antibodies are identified from the sequencing of millions of animal or human B cells after immunological challenge.

Using sequence information as a universal reference system, researchers will be able to choose the binding reagent best suited to their requirements and use them in a standardized way (see 'Reliable binding reagents for all').

Various classes of binding reagents are being developed in addition to antibodies, including recombinant protein scaffolds — proteins with artificially introduced binding surfaces[9] — and nucleic-acid-based binding reagents[10]. Some alternative reagents are easier to use and manufacture than antibodies.

MARKET FORCES

In our view, producing recombinant versions of the antibodies commonly used in research could be commercially profitable — even when the sequence information is publicly available. The absence of these reagents in the research-antibody marketplace stems mostly from economic considerations, rather than technical challenges. Most commercial producers have simply been attracted to the more lucrative therapeutics market.

Production costs for recombinant binding reagents — currently similar to those for monoclonals — should decrease as technologies improve, demand increases and processes become automated. Although researchers could, in principle, make the recombinant reagents themselves if the DNA sequences are freely available, most will prefer the convenience of receiving a quality-controlled product from a supplier. On this basis, two years ago, a UK start-up called Absolute Antibody began selling sequenced monoclonal antibody reagents produced recombinantly. Many of the sequences are already in the public domain.
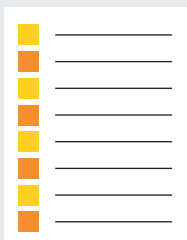
That said, market forces alone are not going to improve the quality of protein-binding reagents — as the past 30 years have shown. Of the nearly 2 million antibodies listed in the database CiteAb, most are used too rarely in research to constitute an appealing commercial prospect.

Achieving a wholesale transition to characterized, recombinant protein-binding reagents will require the public funding agencies of the world's largest economies — North America, Europe and Asia — to make a major investment in technology development. This
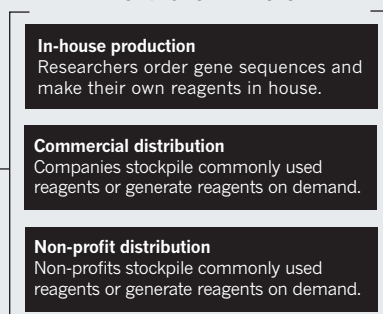
RELIABLE BINDING REAGENTS FOR ALL
Making the sequences of all binding reagents freely available would give researchers and suppliers a universal reference system.
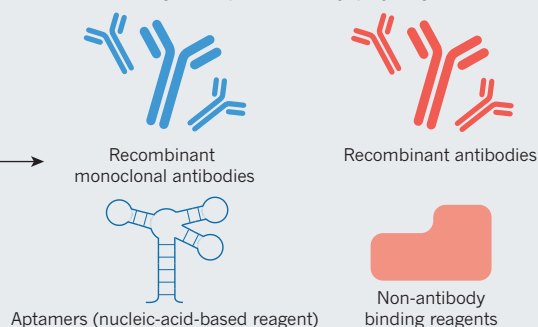

REAGENT SEQUENCES IN DATABASE
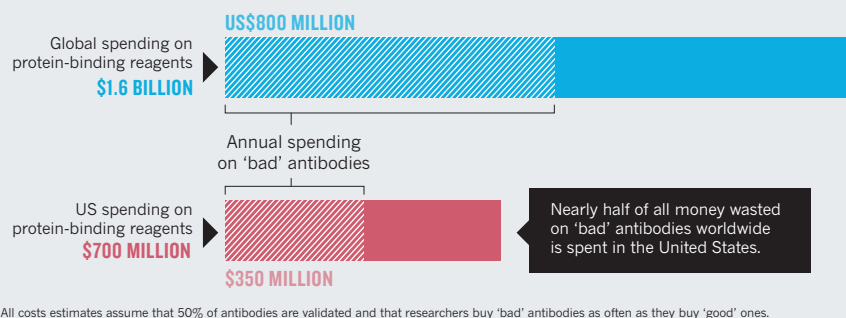
DISTRIBUTION METHODS

In-house production
Researchers order gene sequences and make their own reagents in house.

Commercial distribution
Companies stockpile commonly used reagents or generate reagents on demand.

Non-profit distribution
Non-profits stockpile commonly used reagents or generate reagents on demand.

STANDARDIZED BINDING REAGENTS

Recombinant monoclonal antibodies

Recombinant antibodies

Aptamers (nucleic-acid-based reagent)

Non-antibody binding reagents

MONEY DOWN THE DRAIN

The use of poorly characterized and ill-defined antibodies wastes materials, researcher time and money.

Global spending on protein-binding reagents
$1.6 BILLION

US$800 MILLION

Annual spending on 'bad' antibodies

US spending on protein-binding reagents
$700 MILLION

$350 MILLION

Nearly half of all money wasted on 'bad' antibodies worldwide is spent in the United States.

All costs estimates assume that 50% of antibodies are validated and that researchers buy 'bad' antibodies as often as they buy 'good' ones.

would reduce costs and increase efficiency in the key bottlenecks: the production of reagent targets, the selection of binding reagents and their downstream characterization. Five-year pilot programmes initiated in 2010 — the Protein Capture Reagents Program of the US National Institutes for Health (NIH) and Affinomics by the European Union (EU)— have indicated that recombinant technologies could be scaled up. These programmes should be expanded and investment in them sustained for at least a decade.

We estimate that using current technology, roughly $1 billion would be required to generate characterized recombinant binding reagents to target the primary products of all 20,000 human genes. This is probably less than what is wasted worldwide on bad reagents in two years (see 'Money down the drain'), and would be easily recouped over the long term thanks to more reproducible data.

However, we do not advocate the stockpiling of reagents against targets of minimal interest. Instead, efficient pipelines should be developed to generate any high-quality binding reagent on demand. One possibility is that centres funded by public and private funds would focus on target production, selection, characterization and the publishing of sequence information based on requests from users, whereas commercial companies would specialize in producing and distributing the reagents. An instructive analogue is the Structural Genomics Consortium. This partnership has over the past decade used public and private funds to generate effective production lines for human proteins, deployed by academia and pharmaceutical companies.

As a first step, we ask the scientific leadership of the NIH and the EU to convene academic users, technology developers, biotech companies, funding agencies and publishers, and establish a realistic timetable for the transition to these high-quality binding reagents.

Making sequenced well-characterized reagents is alone unlikely to change the behaviour of researchers. One possible outcome of such a meeting could be that publishers and funding agencies should mandate that in, say, five to ten years time, and contingent on appropriate investment, all binding reagents in published papers are recombinant and defined at the sequence level. This would mirror the requirements for the past few decades that gene sequences and coordinates for new protein structures be deposited and made publicly available.

If these steps are taken, scientists will not want to use unsequenced binding reagents, and the absence of sequencing information will lead to market disadvantages for vendors. The uncharacterized, unsequenced research antibody will become obsolete. ∎

Andrew Bradbury *is a group leader and scientist in the Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.* **Andreas Plückthun** *is professor and head of the Biochemistry Department of the University of Zurich, Switzerland.*
e-mail: amb@lanl.gov

1. Collins, F. S. & Tabak, L. A. *Nature* **505,** 612–613 (2014).
2. Bordeaux, J. *et al. BioTechniques* **48,** 197–209 (2010).
3. Berglund, L. *et al. Mol. Cell Proteomics* **7,** 2019–2027 (2008).
4. Slaastad, H. *et al. Proteomics* **11,** 4578–4582 (2011).
5. Begley, C. G. & Ellis, L. M. *Nature* **483,** 531–533 (2012).
6. Köhler, G. & Milstein, C. *Nature* **256,** 495–497 (1975).
7. Vasilevsky, N. A. *et al. PeerJ* **1,** e148 (2013).
8. Voskuil, J. L. A. *F1000Research* **3,** 232 (2014).
9. Binz, H. K., Amstutz, P. & Pluckthun, A. *Nat. Biotechnol.* **23,** 1257–1268 (2005).
10. Gold, L. *et al. Cold Spring Harb. Perspect. Biol.* **4,** a003582 (2012).

Start research on climate engineering

Safe, small-scale experiments build trust and road-test governance, argue **Jane C. S. Long**, **Frank Loy** and **M. Granger Morgan**.

Climate engineering — cooling Earth intentionally by modifying its radiation balance — worries many people. We know little about the effectiveness of these technologies or their side effects. The unintended consequences could be profound. One country's interventions will affect others and could distract from climate-change mitigation efforts, and there is no international mechanism for regulating such deployments. These are legitimate concerns.

But interventions may need to be considered in the future. The 2013 report of the Intergovernmental Panel on Climate Change suggested that even if the world almost eliminates greenhouse-gas emissions by mid-century, decades of climate engineering — such as removing carbon dioxide from the atmosphere or injecting reflective particles into the stratosphere — might be required to control global temperatures and preserve vulnerable populations and ecosystems[1].

Yet the climate-science community has largely avoided the subject. Government-funded research has been restricted to ▶
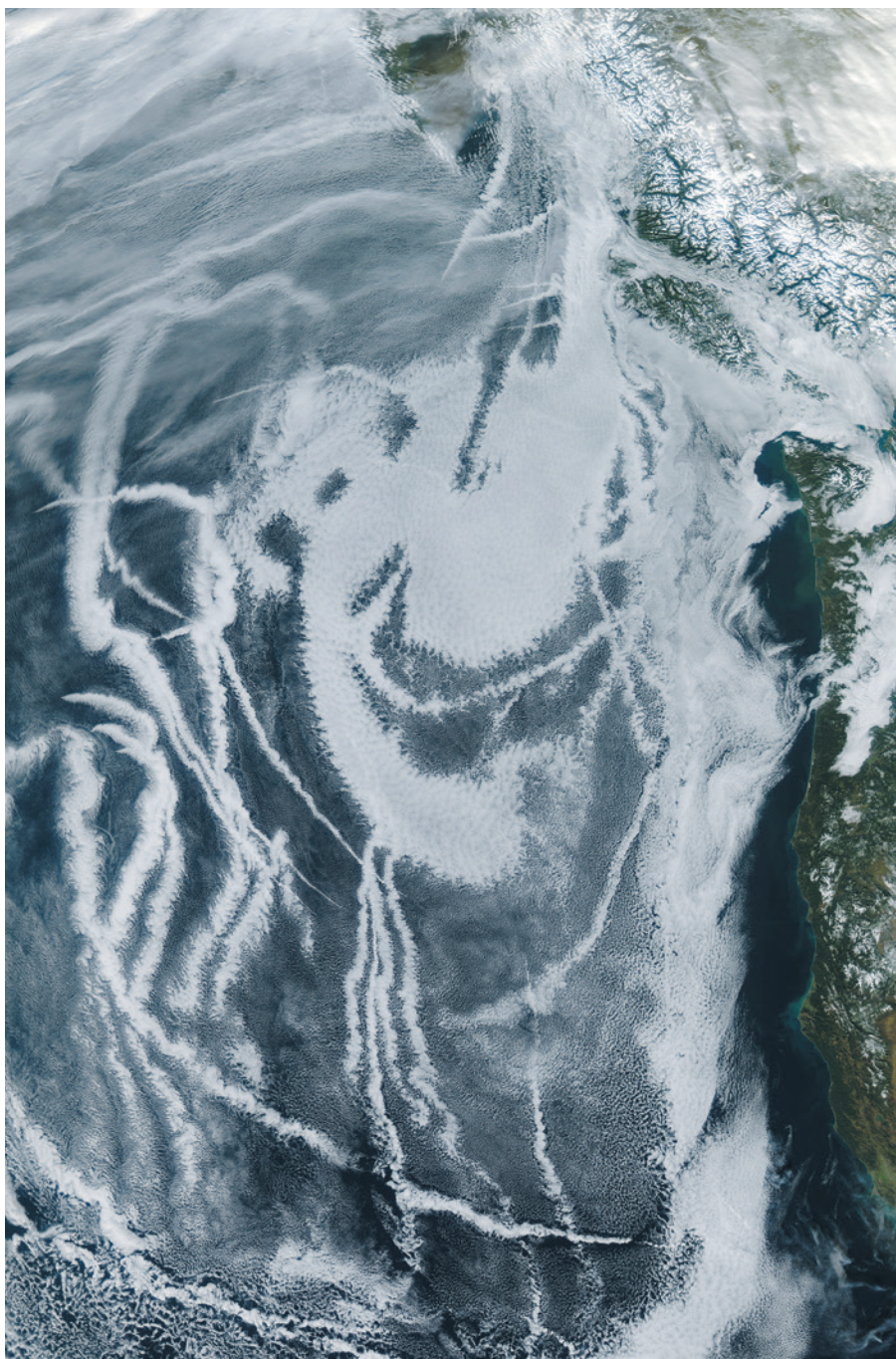
Ship trails (clouds seeded by particles in ship exhaust) off North America on 20 January 2013.

▶ modelling and social-science investigations. The few outdoor experiments that have tested concepts were either funded privately or performed as pure climate science without making the climate engineering intent clear. Such experiments fail to ensure two fundamental principles of good governance of climate-engineering research: transparency and that the research is for the public good.

SOLID UNDERSTANDING

We believe that this laissez-faire approach is risky and imprudent. As the consequences of climate change become starker, public calls for interventions may grow. Governments or companies may try climate engineering to reduce the severe impacts predicted by 2050. Our ignorance of the benefits and problems could become dangerous.

Several reports and institutions have called[2] for climate-engineering research to commence. We agree. We must start now: gaining a solid understanding of any climate-engineering technique will take decades. Small-scale outdoor experiments in particular are needed to provide real-world answers to questions about the efficacy and advisability of climate engineering.

Even the United Nations Convention on Biological Diversity, which has discussed a ban on climate engineering, endorses such experiments (see go.nature.com/vopjwg).

But how should research get started? Should governance be developed before or after early experimentation? Some have called for a moratorium on climate-engineering research until an international governance regime is in place (see go.nature.com/rfx56p). We disagree. A ban would push research underground and towards private funding where risky experiments may proceed ungoverned. Or experiments might be conducted with no more than usual research governance. Neither approach is good.

We argue that governance and experimentation must co-evolve. We call on the US government and others to begin programmes to fund small-scale, low-risk outdoor climate-engineering research and develop a framework for governing it.

START SMALL

Although they do not require approval by an international body, small-scale experiments are an opportunity for international collaboration. Countries that have worked together on small-scale research and participated in developing governance models will be in a better position to agree how to handle risky research should that time ever come.

Opponents of climate-engineering research have claimed that the only useful outdoor research requires perturbing the climate. That is wrong. Many small-scale experiments would have no measurable effect on Earth's climate[3,4]. The physical and chemical processes on which interventions rely need testing and quantification at small scales before any climate impacts are assessed. Experiments that extend up to kilometres in altitude and last days to weeks would leave the global climate unchanged but would increase scientific understanding substantially.

Some useful low-risk experiments have already been identified[5]. Injecting a small amount of sulfur into the stratosphere over several weeks would show how fine particles evolve and affect ozone depletion; spraying salt particles into coastal clouds would assess whether cloud reflectivity can be increased; seeding high-latitude cirrus clouds with artificial ice nuclei would determine whether the clouds can be dissipated and allow more long-wave radiation to escape from Earth.

These small-scale tests look a lot like climate-science experiments, and climate-change science will also benefit from them. Making the intent of the research clear allows governance strategies to be explored. All proposals should address

TESTING CLIMATE ENGINEERING
Checklist for funding research

Value. Climate-engineering experiments should have social as well as scientific benefits, for example by reducing major climate-change uncertainties such as the roles of clouds and aerosols in moderating Earth's energy balance. The research should generate new understanding of the risks, effectiveness and advisability of climate engineering.

Risk. Researchers should evaluate and minimize their proposal's downsides — known, predicted or perceived. Small, short-lived projects raise fewer concerns than large and long ones. Avoid concepts that would permanently alter the environment. Comparing impacts with those of other common activities, such as flying aircraft in the stratosphere, maintains perspective.

Transparency. To maintain trust and ensure that society can learn how to govern climate-engineering research, scientists should conduct experiments openly, facilitate deliberation and oversight, and inform decision-making. Researchers should clearly explain to the public an experiment's scientific context, its intent, method, alternatives, the expected and actual outcomes, and how research questions evolve as a result.

Vested interests. Financial interests and intellectual-property rights may influence research or lead to political pressure that does not serve the public interest. Researchers and institutions could have positive biases about their climate-engineering concepts for professional, intellectual or personal reasons[6].

Governance methods beyond normal peer review are needed to check that conflicts of interest do not bias evaluation. For example, a second team could be asked to confirm or find errors in research done by another.

Legal considerations. Larger-scale research may require environmental regulatory review. For example, the United States may demand an environmental impact assessment or statement under the National Environmental Policy Act, Clean Air Act or Clean Water Act. Experiments that cross national borders must abide by customary international law or United Nations treaties such as the Framework Convention on Climate Change, the Convention on Biological Diversity, or the Convention on the Law of the Sea. If there is foreseeable harm, consent among the affected parties should be determined.

---

five governance considerations: value, risk, transparency, vested interests and legal requirements (see 'Checklist for funding research').

Learning about governance does not follow automatically. The SPICE (Stratospheric Particle Injection for Climate Engineering) experiment proposed for the United Kingdom in 2010 is an example of how not to proceed. This project, which would have simply sprayed water from a hose attached to a tethered balloon, was abandoned after it failed to win public support and when conflict-of-interest issues emerged over a patent application for the system. It aimed to test a mechanism by which climate-altering chemicals could be introduced into the atmosphere to reflect sunlight — but before scientific uncertainties about the effectiveness and advisability of any such interventions had been resolved. Furthermore, little stood to be learned from the experiment, because the hose would have operated at a lower altitude than required for climate engineering. The project became a lightning rod for public concern and was cancelled.

Government agencies and scientists should begin climate-engineering research, learn to govern it and prepare for international collaboration. We recommend the following first steps (developed through discussions at the 2014 Solar Radiation Management Governance Initiative workshop in San Francisco, California).

FIVE STEPS
First, pick a good test case for an outdoor research project. This will establish a track record for dealing with controversy, scrutiny and outreach. The initial experiment should yield valuable scientific insight and be defensible, in that it is brief and poses no significant risk.

Second, clearly identify the research as climate engineering. Obfuscation will violate public trust and obviate co-evolution of governance.

Third, seek broad advice early to identify potential social risks and societal benefits.

"As the consequences of climate change become starker, public calls for interventions may grow."

Such understanding will help when deciding whether to stop or proceed. Think of it as a rehearsal for constructing an advisory body, should the government decide to establish a strategic research programme.

Fourth, discuss climate engineering within the broader context of climate-change strategy. Climate engineering cannot substitute for mitigation or adaptation, but it might (or might not) provide crucial tools in a holistic and strategic plan for dealing with the inevitable impacts of global change.

And fifth, assess the early work and decide whether and how to proceed. What was learned? Do the results render any subsequent approaches untenable or indicate that a modification would be more effective or more advisable? What new scientific issues are identified? What are the next steps? If public concerns are raised, how can engagement be more

effective and useful?

Government agencies must take these steps. To ensure transparency and public trust, outdoor experiments in climate-engineering should be publicly, rather than privately, funded.

We urge researchers to come forward with well-crafted proposals that meet the test-case requirements. Global collaborators should be engaged as a precursor to more formal international cooperation. ■

Jane C. S. Long *is former associate director of Lawrence Livermore National Laboratory in Livermore, California,USA.* **Frank Loy** *is a former United States Under Secretary of State for Global Affairs (1998–2001).* **M. Granger Morgan** *is at Carnegie Mellon University in Pittsburgh, Pennsylvania, USA.*
e-mail: janecslong@gmail.com

1. Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).
2. Bipartisan Policy Center Task Force on Climate Remediation Research. *Geoengineering: A National Strategic Plan for Research on the Potential Effectiveness, Feasibility, and Consequences of Climate Remediation Technologies* (Bipartisan Policy Center, 2011).
3. Parson, E. A. & Keith, D. W. *Science* **339,** 1278–1279 (2013).
4. Morgan, M. G. & Ricke, K. *Cooling the Earth Through Solar Radiation Management: The Need for Research and an Approach to its Governance.* (International Risk Governance Council, 2010).
5. Keith, D. W., Duren, R. & MacMartin, D. G. *Phil. Transact. R. Soc. A.* **372,** 20140175 (2014).
6. Long, J. C. S. & Scott, D. *Issues Sci. Tech.* 45–52 (Spring issue, 2013).

Physicist Bruno Pontecorvo (left) defected to the Soviet Union in 1950.

NUCLEAR PHYSICS

New light on a cold war conundrum

Sharon Weinberger ponders a chronicle claiming that fresh evidence has cracked the 'Pontecorvo affair'.

When the gifted Italian nuclear physicist Bruno Pontecorvo vanished with his family in the late summer of 1950, no one took any notice. It was more than two weeks before friends, family and, most importantly, the intelligence agencies of the United States and Britain — both countries where Pontecorvo had worked — realized that he had gone. And it was another five years before it was confirmed that Pontecorvo had defected to the Soviet Union.

The defection is one of the last great unsolved cold-war science mysteries. Did Pontecorvo, apparently on his way to winning a Nobel prize, flee because he was a devoted communist facing a hostile political climate? Or was he a spy running to his handlers? Physicist Frank Close believes that he has the answer. In *Half-Life*, Close explores what is now known as the Pontecorvo affair. Woven into that chapter in the physicist's life are two love stories: his Swedish wife's love for him, and his love for communism. Neither ended well.

Born in 1913 into a successful Jewish family near Pisa, Pontecorvo pursued a career in physics, first as one of Enrico Fermi's "Via Panisperna" boys, and then leaving for Paris in 1936 to escape the rise of fascism and anti-Semitism. In France, he worked in the lab of Irène and Frédéric Joliot-Curie, and was swept up in the communist movement. Here he met and married his wife, Marianne Nordblom. As the war made its way to France, Pontecorvo and Marianne — now with a child — left for Tulsa, Oklahoma, where he used his expertise in looking at how neutrons interact with materials to create a new method for oil prospecting. But Pontecorvo's politics drew the attention of the FBI.

The federal investigators were not crack spy hunters. FBI agents visited Pontecorvo's Tulsa home in 1942 — when the United States was at war with Italy — and found communist literature in plain view. The couple's second son was named Tito, after the leader of communist Yugoslavia. Short of flying the hammer and sickle, it is hard to see how Pontecorvo could have been more open about his leanings. Yet the FBI merely wrote a report, which languished for several years. In the meantime, Pontecorvo moved to Canada to work on the Manhattan Project at the Chalk River Laboratories, and in 1949 left for the United Kingdom to work for the Atomic Energy Research Establishment at Harwell.

All seemed well until midway through the fateful summer holiday in 1950. The family, by then with three children, suddenly made a dash for Sweden, then Finland; here, Soviet agents secreted them across the border, with Pontecorvo hiding in the boot of a car. That marked the beginning of what Close calls the physicist's "half-life": the end of his Western career and the beginning of his life as Soviet scientist Bruno Maksimovich Pontecorvo.

Too many books are fêted as reading 'like spy novels', but Close's work deserves the accolade. He makes a good circumstantial case for Pontecorvo being a spy. Some of the evidence is relatively convincing, such as the assertion of a former KGB agent who claims that Pontecorvo did spy for the Soviets. On the weaker side, Close cites two pictures of Pontecorvo in Canada — standing with Chalk River scientists — looking away from the camera. Close suggests that Pontecorvo
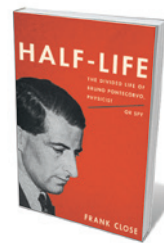
Half-Life: The Divided Life of Bruno Pontecorvo, Physicist or Spy
FRANK CLOSE
Basic/Oneworld: 2015.

was already going to the US–Canadian border to rendezvous with a Soviet agent, and did not want his face recognized. Possibly; or perhaps he looked away by accident.

A bigger problem with this spy narrative is that it lacks a denouement — the dramatic unmasking of a double agent. Close implies that he has finally got the goods in an archival document: a letter from the British Embassy in Washington DC to the director-general of the UK intelligence agency MI5, dated 13 July 1950. It was this, Close writes, that would

> "Too many books are fêted as reading 'like spy novels', but Close's work deserves the accolade."

"lead me to solve the mystery of Bruno Pontecorvo's sudden disappearance". The letter is proof that FBI reports were finally making their way to Europe. Close argues that British double agent Kim Philby, who had access to it, tipped off Moscow; the Soviets then warned Pontecorvo that he was about to be exposed as a spy and arranged for his retreat.

Close's conclusion is in stark contrast to that of historian Simone Turchetti's *The Pontecorvo Affair* (University of Chicago Press, 2012). That book concluded that the Italian physicist was simply a committed communist, whose flight was triggered by a US patent lawsuit that he feared would lead to political persecution. Close's version is more plausible, but both lack definitive proof.

Close is at his best when describing Pontecorvo's work in neutrinos and neutron detection, demonstrating how groundbreaking it was, in spite of later attempts by Western governments to downplay his importance. But Close also occasionally takes liberties for the sake of drama. He describes an interrogation room painted in grey and "mustard", but his citation notes that he bases this scene on his knowledge of Soviet offices, not source materials. Almost nothing is known of Pontecorvo's dealings with the Soviet government. It is one thing to evoke colour, quite another to paint with the entire palette.

Half-Life reveals the real victim as Marianne, whose already shaky mental health deteriorated precipitously in the Soviet Union. Cut off from family and friends, she spiralled into depression and repeatedly entered psychiatric institutions, while Pontecorvo took on a mistress. Pontecorvo's love affair with communism did not end any better. He conceded in 1992, after the fall of the Soviet Union, that his dreams of a communist utopia were an illusion. "I was," he told a reporter, "a cretin." ∎

Sharon Weinberger *is a journalist specializing in national-security issues. She is currently writing a history of the US Defense Advanced Research Projects Agency. e-mail: sharonweinberger@gmail.com*

Books in brief
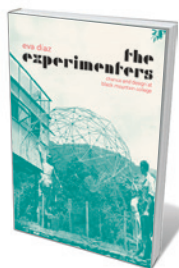
The Experimenters: Chance and Design at Black Mountain College
Eva Díaz UNIVERSITY OF CHICAGO PRESS (2015)
What links systems theorist and architect R. Buckminster Fuller with artistic innovators such as Josef Albers and John Cage? The answer is Black Mountain College, North Carolina. From 1933 to 1957, in this unaccredited institution in Appalachia, they and other "artist-scientists" created an iconic lab for experimental research in the arts. As art historian Eva Díaz reveals in this engrossing study, their explorations in materials, form, chance and indeterminacy were never less than electrifying. Her sympathetic portrait of Fuller as a utopian saving the world through geodesic geometry is particularly assured.
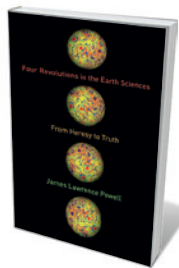
Four Revolutions in the Earth Sciences: From Heresy to Truth
James Lawrence Powell COLUMBIA UNIVERSITY PRESS (2014)
Deep time, continental drift, meteorite impact and climate change: each of these twentieth-century geoscientific discoveries was once viewed as heretical. So reminds geologist James Powell in this exemplary treatise on scientific progress. He traces the evolution of each landmark finding through the work of the dogged researchers who proved it, step by step. Many fights were hard-won, as shown in the efforts of Gene Shoemaker, Luis and Walter Alvarez, Robin Canup and others who established connections between meteorite impact, the birth of the Moon and the extinction of the dinosaurs.

Finding Zero: A Mathematician's Odyssey to Uncover the Origins of Numbers
Amir D. Aczel PALGRAVE MACMILLAN (2015)
Mathematician Amir Aczel was obsessed from childhood with the origins of numerals. This bracing mathematical detective story reveals how he cracked the puzzle: by homing in on zero. Close readings of classical texts convinced him that this subtle concept emanated from the East. He treks through the findings of archaeological scholar George Cœdès, the surprising nexus of sex and mathematics, and much of southeast Asia before hitting pay dirt with a seventh-century artefact in a dusty Cambodian shed.

A Scientist in Wonderland: A Memoir of Searching for Truth and Finding Trouble
Edzard Ernst IMPRINT ACADEMIC (2015)
During his 1993–2011 tenure as the world's first chair in complementary medicine (at the University of Exeter, UK), Edzard Ernst scrutinized alternative medical treatments, turning up false claims and sparking a furore among enthusiasts. As he shows in this ferociously frank autobiography, his early career was as dramatic — during a stint as chair of rehabilitation medicine at the University of Vienna, he uncovered the institution's historical involvement in medical experiments under the Third Reich. A clarion call for medical ethics.

Pioneers of Neurobiology: My Brilliant Eccentric Heroes
John G. Nicholls SINAUER ASSOCIATES (2015)
This scientific memoir by neurobiologist John Nicholls takes the form of short biographical sketches focusing on the eccentricities of key people he has worked with or encountered, from Nobel prizewinners to lab technicians. It is quite a list, including Stephen Kuffler, Bernard Katz, Rita Levi-Montalcini, Gunther Stent and James Watson. Nicholls's gently amusing anecdotes shed light on the developing environment of molecular neurobiology, mostly in Europe and the United States, since the 1950s. **Barbara Kiser**

Correspondence

More than 70 ways to show resilience

Ruben Dahm and his colleagues call for greater "flood resilience" in delta cities (*Nature* **516,** 329; 2014). But achieving resilience depends on what we mean by it: there are more than 70 definitions in the scientific literature.

These definitions vary between two extremes, with most trying to achieve a balance between the two (M. de Bruijne *et al.* in *Designing Resilience* 13–32; Univ. Pittsburgh Press, 2010). Dahm *et al.* implicitly define resilience as the ability of a system to bounce back after stress, as do many politicians and the World Economic Forum in Geneva, Switzerland.

At the other extreme, resilience is seen as "the capacity of social–ecological systems to adapt or transform in response to unfamiliar, unexpected and extreme shocks" as proposed by ecologist Stephen Carpenter and economics Nobel laureate Kenneth Arrow, among others (S. R. Carpenter *et al. Sustainability* **4,** 3248–3259; 2012).

To recover, to adapt or to invest in both possibilities? All make sense in the right context. We might want a city to recover from flooding, for example, or the world to adapt to the effects of climate change.

Long-term policies to promote either recovery or adaptation, or prepare for both, are likely to be very different. We must have a clear definition from the outset.
Len Fisher *University of Bristol, UK.*
len.fisher@bristol.ac.uk

Labs leak staff under French law

Ironically, a well-intentioned law enacted in France almost three years ago to protect the poorly qualified is preventing young researchers from completing their postdoctoral training. In my view, research institutions should be exempted from its requirements.

Called the Sauvadet law, it stipulates that a worker must be appointed to a permanent position after six years of short-term contracts in the public sector. Since the law came into effect in March 2012 (and retrospectively), publicly funded research institutions — including INSERM and the CNRS — have limited the number of postdocs becoming eligible for tenure by not renewing contracts if a postdoc has already worked there for three years (see go.nature. com/vki6fq and go.nature.com/ navteh; both in French).

Young researchers are therefore being forced to complete their training abroad. The law also means that invaluable laboratory engineers and senior technicians can no longer be retained under short-term contracts if a permanent position is not available. The nation urgently needs to put countermeasures in place, or risk losing crucial lab staff indefinitely.
Juan Iovanna *INSERM UMR1068, CNRS UMR7258, Aix-Marseille University and Institute Paoli-Calmettes, Marseille, France.*
juan.iovanna@inserm.fr

Replace 'pathogens' with 'perceptogens'

We argue for a more sweeping reappraisal of the term pathogen than Arturo Casadevall and Liise-anne Pirofski propose (*Nature* **516,** 165–166; 2014). This should take in not just microbes, but the wider 'exposome' and recent discoveries in infection and immunity research.

A term is needed that encompasses sequences from the environment — intrinsic or extrinsic — that impart pathogenic or benign information to eukaryotic immune receptors. For example, T-cell receptors that recognize autoantigen and microbial sequences can be triggered by related peptide sequences from diverse sources in the environment (M. E. Birnbaum *et al. Cell* **157,** 1073–1087; 2014).

Substituting 'microbial immunogen' for 'pathogen' would not account for microbiota sequences that instruct immune development rather than elicit protective immunity. We suggest instead the term perceptogen (microbial or environmental) to cover protein sequences that affect the body's range of reactions after perception by its immune receptors.

As the writer Aldous Huxley remarked: "There are things known and there are things unknown and in between are the doors of perception."
Danny Altmann, Rosemary Boyton *Imperial College London, UK.*
d.altmann@imperial.ac.uk

Learning from Typhoon Haiyan

In our view, the communication of disaster risk during Typhoon Haiyan, which struck the Philippines in November 2013, could have been better.

The typhoon was one of the strongest tropical storms ever to make landfall, registering category 5 on the Saffir–Simpson scale. Despite forecasts of winds of more than 300 kilometres per hour and a predicted 7-metre storm surge, the city of Tacloban was caught underprepared: thousands died from the inundation.

The storm surge was predicted in a report by the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) that was sent to local agencies and communities. Unfortunately, it was simply a line at the end of a routine weather bulletin. It was apparently not otherwise highlighted, elaborated on or, in our opinion, in any way explained in order to transmit its urgency to key agencies and the public.

After interviewing agency personnel, we concluded that a well-intended adherence to routine and pro forma communication could have been at play. Feedback loops for conveying tacit information (for example, the implications of modelling outputs) seem to have been inadequate.

PAGASA's Tacloban team stayed in its single-storey coastal office, which was demolished by the storm surge, claiming a team member's life.

Many other factors influenced the impact of Haiyan, but this example indicates that routines need to adapt to deal with extreme events that lie beyond personal and institutional memory.
Raul Lejano *New York University, New York, USA.*
Joyce Melcar Tan, Meriwether Wilson *University of Edinburgh, UK.*
lejano@nyu.edu

Build neuroscience capacity in Africa

The non-profit organization TReND is funding a neuroscience training initiative so that Africa's scientists can join this rapidly evolving research field (see www.trendinafrica.org).

TReND (for 'Teaching and Research in Neuroscience for Development') is run by volunteer researchers at several universities worldwide. It organizes outreach courses and workshops for young African scientists on how to conduct quality, affordable neuroscience research in resource-limited settings. In 2010–14, more than 1,000 African students took part in TReND programmes.

The organization provides students with used lab equipment from universities, hospitals and companies in developed countries, as well as open-source software and hardware.

We call on industry and governments for more investment in such activities to improve science education and promote economic development in Africa.
Fanuel Muindi, Joseph Keller *Massachusetts Institute of Technology, Cambridge, USA.*
fmuindi@mit.edu

Mary F. Lyon
(1925–2014)

Grande dame of mouse genetics.

In 1961, Mary Frances Lyon proposed in *Nature* that one of the two X chromosomes in every cell of female mammals is inactivated. This, she argued, occurs to prevent XX female cells from expressing twice as many X-linked gene products as XY male cells.

Lyon's X-chromosome inactivation hypothesis had profound implications for clinical genetics and developmental biology. For instance, it helped researchers to elucidate the genetic basis of many X-linked diseases, such as Duchenne muscular dystrophy. It also led, 30 years later, to the discovery of the *Xist* gene, which helped to spawn a whole field of research on the role of long non-coding RNA molecules in regulating gene expression. The non-coding Xist RNA is the master switch that turns off the X chromosome. It also explains some everyday phenomena, such as the patchy colouring of tortoiseshell cats.

Lyon died aged 89, on Christmas Day 2014, after drinking a glass of sherry, eating her Christmas lunch and settling down in her favourite chair for a nap. She was born in Norwich, UK, in 1925, the eldest of three children of a teacher mother and civil-servant father. An inspirational teacher at her grammar school sparked her interest in biology and in 1943 — when women were awarded only 'titular' degrees — she went to Girton College, University of Cambridge, to study zoology.

In 1946, Lyon started her PhD at Cambridge with the geneticist R. A. Fisher in the emerging field of mouse genetics. In 1948, in pursuit of better histology facilities to complete her doctoral studies, Lyon moved to the Institute of Animal Genetics at the University of Edinburgh, at the time headed by the embryologist C. H. Waddington, another figure whose work profoundly influenced her.

Lyon stayed in Edinburgh to work on a project funded by the Medical Research Council (MRC) to study mutagenesis in mice under the geneticist Toby Carter: in the 1940s there was widespread concern about the possibility of atomic-weapons testing causing mutations. In 1955, Carter transferred his group to the MRC Radiobiology Unit at Harwell, UK, where there was more space for mouse breeding. It was during this period that Lyon observed the patchy coats of female mice carrying X-linked coat-colour mutations. This, coupled with the knowledge that female mice carrying a single X chromosome are

COURTESY MRC HARWELL/MARY LYON

viable, led her to propose the hypothesis of X-chromosome inactivation.

Apart from a short sabbatical in Cambridge, Lyon remained at Harwell for the rest of her career. From 1962 she was head of the genetics division, which became an independent Mammalian Genetics Unit in 1995.

Despite some initial tussles with the MRC over the amount of time and resources to devote to 'ancillary projects' in developmental genetics — rather than to establishing the hazards of radiation and other mutagenic agents — Lyon managed to pursue both.

Throughout her six decades of work on mice, her favourite chromosome, aside from the X, was 17. Chromosome 17 encodes the t-complex, a genetic anomaly found in certain wild mice that gives rise to different 't-haplotypes', which consist of different DNA rearrangements. Certain t-haplotypes are preferentially transmitted by males to their offspring; mice carrying two copies of the same t-haplotype are either not viable or sterile. By carrying out a series of clever genetic crosses, Lyon worked out what was going on. This work made a major contribution to the understanding of phenomena such as non-Mendelian inheritance (the abnormal segregation of chromosome pairs from the expected one-to-one ratio) and the effect that inversions — when a segment of a chromosome is reversed — have on

suppressing chromosomal recombination.

Lyon was a central figure in twentieth-century mouse genetics. She laid the intellectual foundations and developed the genetic tools for the use of mice as model organisms in molecular medicine, cell and developmental biology and in deciphering the function of the human genome. Lyon was editor of *Mouse News Letter* from 1956 to 1970, a publication that had a key role in establishing a mouse-focused research community in the pre-Internet age. She also helped to develop a common language for the field by chairing the Committee on Standardised Genetic Nomenclature for Mice from 1975 to 1990. Her pivotal contribution was recognized by the naming of the Mary Lyon Centre, an international facility for mouse-genetic resources, opened at Harwell in 2004, and by the creation of the Mary Lyon Medal by the UK Genetics Society in 2014.

Because everything Mary said was so carefully thought through, she could be difficult to talk to: on the phone, it was easy to think you had been cut off. She did not suffer fools gladly, but was a great supporter of the bright young scientist, often eschewing authorship of publications to enhance the profile of junior collaborators. She was intellectually rigorous but not dictatorial. When I began my PhD with her in 1977, she gave me a handful of papers, showed me the genetic tools — mice carrying the various mutations and chromosomal rearrangements — and said, "do something on X-inactivation". That degree of academic freedom was exhilarating, coupled as it was with the safety net of robust critique.

Among numerous other honours, Mary was a foreign associate of the US National Academy of Sciences and was the 28th eighth woman to be elected a fellow of the Royal Society in London in 1973. She might have been elected sooner had leading geneticist Hans Grüneberg not initially disbelieved the Lyon hypothesis. It is perhaps surprising that Mary did not receive any establishment honours, but bureaucracy, politics and networking were alien to her.

Her first love was mice, although she always had a cat — a tortoiseshell, of course. ■

Sohaila Rastan *is executive director of biomedical research at Action on Hearing Loss in London, UK. She was Mary Lyon's second PhD student at the MRC Radiobiology Unit in Harwell, UK.*
e-mail: sohaila@rastan.fsnet.co.uk

CHEMICAL BIOLOGY

How to minimalize antibodies

The success of antibodies as pharmaceuticals has triggered interest in crafting much smaller mimics. A crucial step forward has been taken with the chemical synthesis of small molecules that recruit immune cells to attack cancer cells.

CHRISTOPH RADER

Over the past two decades, a wave of biological medical products, or biologics, dubbed monoclonal antibodies (mAbs) has conquered the pharmaceutical armamentarium. There are more than 30 mAbs marketed for treating cancer, autoimmune diseases and other serious medical conditions, and a similar number are in late-stage clinical trials[1]. At least five mAbs have each garnered more than US$5 billion in annual revenue. Small molecules that mimic the pharmacological properties of mAbs therefore have the potential to become highly competitive drugs. Writing in the *Journal of the American Chemical Society*, McEnaney *et al.*[2] provide evidence that this is an achievable goal.

The success of mAbs as pharmaceuticals is remarkable, given their size, composition and heterogeneity (mAbs are populations of similar, but not identical, molecules). They are more than 100 times larger than conventional small-molecule drugs manufactured using chemical synthesis, and therefore require more-expensive and less-precise biological synthesis.

Tripartite Y-shaped antibodies evolved as a cornerstone of the vertebrate immune system. They can selectively and tightly bind to foreign molecules with their two Fab regions (targeting functions) and recruit components of the host immune system with their Fc region (effector functions). Furthermore, the Fc region mediates recycling of the antibody molecule, resulting in its retention in the blood (giving it a prolonged circulatory half-life). In the treatment of cancer, mAbs use their tripartite architecture to bring immune cells readied for the kill in close proximity to cancer cells (Fig. 1a).

These three principal features of antibody molecules bestow pharmacological properties on mAbs that are unmatched by small molecules. But the complexity of mAbs has prevented generic versions of branded drugs from being produced by competitor companies, and slowed the production of similar versions — providing a strong investment incentive for pharmaceutical companies, but potentially driving up health-care costs. By contrast, small molecules that mimic mAbs
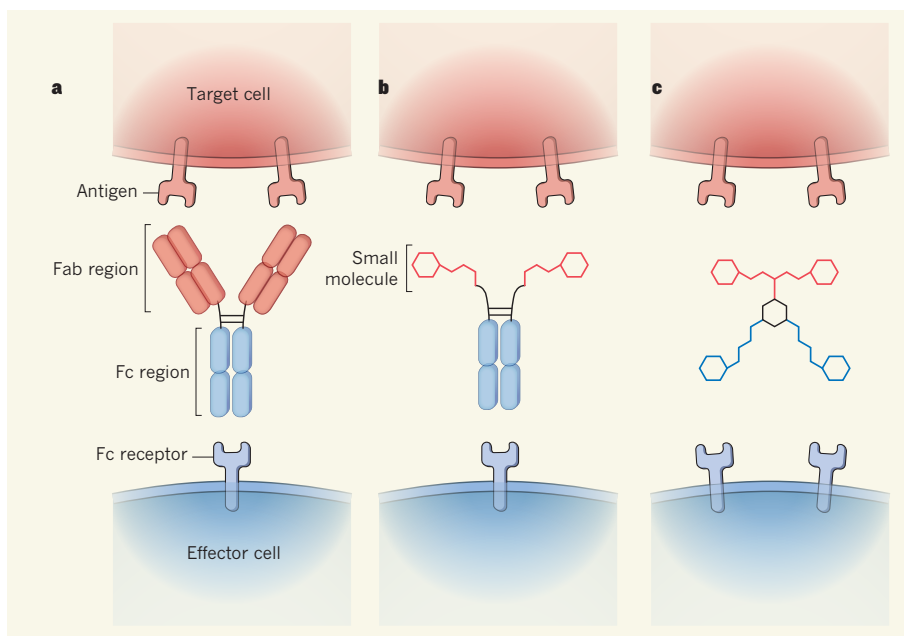

Figure 1 | Bridging target and effector cells. a, Tripartite antibody molecules bridge target and effector cells by simultaneously engaging two structures (antigens) on the target-cell surface with their Fab regions and one effector-cell surface receptor with their Fc region. **b**, Small molecules that mimic Fab regions can be docked to whole antibodies (not shown) or Fc regions (depicted) to produce molecules that have the pharmacological properties of antibodies. **c**, McEnaney *et al.*[2] now report a quadripartite synthetic antibody that is 20 times smaller than naturally occurring antibody molecules and readily generated by chemical synthesis. Cartoon versions of the antibody and the small molecules are approximations of the actual molecular structures and not drawn to scale.

would have lower manufacturing costs and enable competition from generics. And unlike mAbs, which often trigger immune responses in patients, small molecules are not immunogenic. Moreover, they can penetrate tissues and cells more efficiently, can reach buried sites on target molecules that are inaccessible to mAbs, can be given orally and have a longer shelf life.

The discovery and development of peptides, peptidomimetics and other small molecules that have a specificity and affinity for biological targets comparable to those of mAbs have been key in efforts to replace mAbs by small molecules. In order for them also to have the effector functions and prolonged circulatory half-life of mAbs, these small molecules were designed to dock to antibodies either *in vitro*, yielding chemically programmed antibodies[3], or *in vivo*, producing antibody-recruiting molecules[4]. Although the resulting hybrid

molecules (Fig. 1b) combine several of the advantages of mAbs and small molecules, their biological component still restricts the scope of their chemical component.

A provocative question has therefore been whether small molecules can copy both the Fab- and Fc-mediated pharmacological properties of mAbs. This is theoretically possible if Fab-mimicking small molecules can be fused to other small molecules that bind to Fc receptors. McEnaney *et al.* now deliver this missing link. They used chemical synthesis to combine a known Fab-mimicking small molecule that binds to a cell-surface receptor on prostate-cancer cells with a known Fc-mimicking cyclic peptide that selectively binds to an Fc receptor called CD64 on immune cells. The resulting compound mimics two of the three principal natural features of antibodies.

Bridging two cell-surface receptors is a formidable task for a small molecule. It

requires a linker between the Fab- and Fc-mimicking components that has sufficient length, solubility and rigidity. The authors used computer modelling to predict that less than one-third of the naturally occurring distance between Fab and Fc regions in an antibody is required to simultaneously engage receptors on two different cells, and they used this information to design their linker. They also found that two copies of each Fab- and Fc-mimicking component are essential for efficiently mediating targeting and effector functions *in vitro*. The result is a quadripartite molecule (Fig. 1c) that resembles tripartite antibodies with respect to composition and function, but which is 20 times smaller, homogeneous (all the molecules are the same) and readily generated by chemical synthesis. Encouragingly, McEnaney *et al.* observed that this compound induces immune cells to engulf and ingest prostate-cancer cells *in vitro*.

Although *in vivo* validation studies have still to be performed, it seems that McEnaney *et al.* have taken a pivotal step towards obtaining antibody-mimicking small molecules that avoid some of the liabilities of biologics. But there is more work to be done. First, the quadripartite molecule is about 7 kilodaltons in size — substantially smaller than antibodies, but still larger than conventional small-molecule drugs (less than 1 kDa), which limits most of the potential advantages discussed earlier. However, the molecular weight can conceivably be cut further by replacing the relatively large Fc-mimicking cyclic peptide with a peptidomimetic or other small molecule. Then again, an intermediate-sized synthetic antibody mimic might be a good thing, because unrestrained access of small molecules to intra- and extracellular nooks and crannies could make their activity and toxicity profiles unpredictable.

Second, the Fc-mimicking component of the synthetic antibody mimic binds to only one kind of Fc receptor (CD64), whereas natural antibodies and mAbs engage other Fc receptors, including CD16 and CD32, on a variety of functionally different effector cells. Moreover, the Fc region of antibodies triggers activation of the complement cascade, which is an additional mechanism of target-cell destruction, and it also mediates prolonged circulatory half-life — effects that generally have been difficult to produce with small molecules. Even so, synthetic antibody mimics that engage only one kind of Fc receptor might allow effector functions to be fine-tuned. The modular, versatile design of McEnaney and co-workers' molecules will also allow their properties to be tailored through chemical synthesis, for example to include a peptide or peptidomimetic that is retained in the blood by binding to circulating albumin protein[5].

Third, Fab-mimicking small molecules are still limited in scope when compared with mAbs, which can be generated and evolved to bind to almost any cell-surface receptor selectively and tightly[3]. Nonetheless, our ability to generate and screen large chemical libraries, which are structurally much more diverse than biological libraries, has afforded access to an increasing number of small molecules that can compete with mAbs in terms of specificity and affinity[6].

This compound induces immune cells to engulf and ingest prostate-cancer cells in vitro.

Meanwhile, proponents of biologics are not sitting idle. Antibody engineers have generated a large variety of antibody molecules that have improved targeting and effector functions. For example, a new class of 'bispecific' antibody[7] can recruit and activate T cells, which are particularly potent effector cells that cannot be directly engaged by natural antibodies and mAbs. Although not as miniaturized as synthetic antibody mimics, these bispecific antibodies are three times smaller than mAbs and can be clinically potent, safe and profitable, as demonstrated by the recently marketed anticancer drug blinatumomab[7]. Intriguingly, however, synthetic antibody mimics might be better copies of these T-cell-engaging biologics than of conventional mAbs, because the biologics bind to just one kind of effector-cell receptor (CD3) and do not need prolonged circulatory half-lives. All things considered, synthetic antibody mimics have the potential to become a new class of pharmaceutical. ∎

Christoph Rader *is in the Departments of Cancer Biology and Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida 33458, USA.*
e-mail: crader@scripps.edu

1. Reichert, J. M. *mAbs* **7**, 1–8 (2015).
2. McEnaney, P. J. *et al. J. Am. Chem. Soc.* **136**, 18034–18043 (2014).
3. Rader, C. *Trends Biotechnol.* **32**, 186–197 (2014).
4. McEnaney, P. J., Parker, C. G., Zhang, A. X. & Spiegel, D. A. *ACS Chem. Biol.* **7**, 1139–1151 (2012).
5. Dennis, M. S. *et al. J. Biol. Chem.* **277**, 35035–35043 (2002).
6. Kodadek, T. *Curr. Opin. Chem. Biol.* **14**, 713–720 (2010).
7. Frankel, S. R. & Baeuerle, P. A. *Curr. Opin. Chem. Biol.* **17**, 385–392 (2013).

The slippery base of a tectonic plate

High-resolution imaging of the base of the Pacific plate as it descends beneath New Zealand discloses a 10-kilometre-thick channel that decouples the plate from underlying upper mantle. See Letter p.85

CATHERINE A. RYCHERT

In the theory of plate tectonics, the outer shell of the Earth, known as the lithosphere, consists of several rigid plates, which move relative to each other over the weaker, flowing asthenosphere. The bottom of the lithosphere, the lithosphere–asthenosphere boundary (LAB), is fundamental to our understanding of how plate tectonics works, although an exact understanding of the mechanism that gives the plates their rigidity and defines their thickness remains elusive and widely debated. On page 85 of this issue, Stern *et al.*[1] describe how they have used reflected seismic waves generated by explosive sources in steel-cased boreholes to image the Pacific plate as it descends beneath New Zealand. They find a LAB that is less than 1 kilometre thick at the top of a 10-km-thick channel, in which slow seismic velocities may require the presence of water or melt (Fig. 1). The authors suggest that the thin channel decouples the lithosphere from the asthenosphere and allows plate tectonics to take place. The existence of such a localized channel probably has implications for the driving forces of plate tectonics and mantle dynamics.

Plate tectonics has been a fundamental tenet of Earth science for almost 50 years. It is the foundation of modern Earth science, and provides a framework for our understanding of the formation of the continents and ocean basins, and the evolution of the planet. However, questions remain, such as, where is the base of a plate and what makes a plate 'plate-like'? There are many proxies used to estimate the depth and nature of the base of tectonic plates, but so far no consensus has been reached. The transition from the rigid lithosphere to the flowing asthenosphere has classically been defined by temperature. Temperature has a large effect on the viscosity of rocks — their ability to flow.

If temperature alone were the sole mechanism governing the definition of the plate,

then we would expect a gradual transition from the lithosphere to the asthenosphere. However, in the past decade the high resolution provided by imaging techniques based on measurements of seismic waves scattered from the base of the lithosphere has revealed that the transition from the lithosphere to the asthenosphere is sharp[2]. This suggests that another mechanism such as the presence of water or melt must exist in the asthenosphere, which would weaken it[3,4], and thus necessarily define the LAB[2].

Stern et al. used data from explosive-source seismic waves that travelled deep into the Earth and were reflected back to the surface, where they were recorded by seismometers in New Zealand. The seismic waves vibrated at high frequency, allowing the authors to image the low-seismic-velocity channel and also to deduce the thickness of the LAB. The deduced LAB thickness is one of the tightest constraints so far on the transition from the lithosphere to the asthenosphere. Similarly significant is the reported thickness of the 120-million-year-old Pacific plate at 73 ± 1 km, much thinner than predicted by the classic thermal model of conductive cooling of the oceanic lithosphere with time[5]. Finally, an increase in seismic-wave velocity about 10 km deeper and parallel to the LAB is interpreted as the base of a decoupling channel.

If the deduced LAB represents the base of the plate, the plate's thinness may explain the enigmatic observed lack of subsidence for sea floor that is more than 70 million years old[6]. However, whether or not it is the plate base depends on the mechanism responsible for the authors' observations. A seismic-velocity discontinuity imaged beneath the Pacific plate at similar depths is sometimes interpreted as anisotropic fabric[7–11] — the directional dependence of seismic-wave velocity. A purely anisotropic interpretation for the observed seismic-velocity discontinuity would not necessarily equate it with the LAB because anisotropic fabric could be frozen into the plate from a previous episode of deformation. Although not impossible here, an exotic anisotropic fabric would probably be required that may not be consistent with typical notions of horizontal fast directions in the lithosphere (see, for example, ref. 12).

Therefore, the authors are left with the hypothesis that water or melt is present in the channel, which would weaken the mantle[3,4] and define the base of the plate[2]. An increase in hydration[13] with depth could be related to the shallow dehydration that occurs during plate formation at a mid-ocean ridge[14] (Fig. 1), whereas melt could be caused by complex mantle flow from subduction tectonics and/or melt ponding[15]. Further investigation is needed to find the origin of any existing melt, because normal oceanic lithosphere is predicted to be cold at a depth of 73 km and so is not necessarily predicted
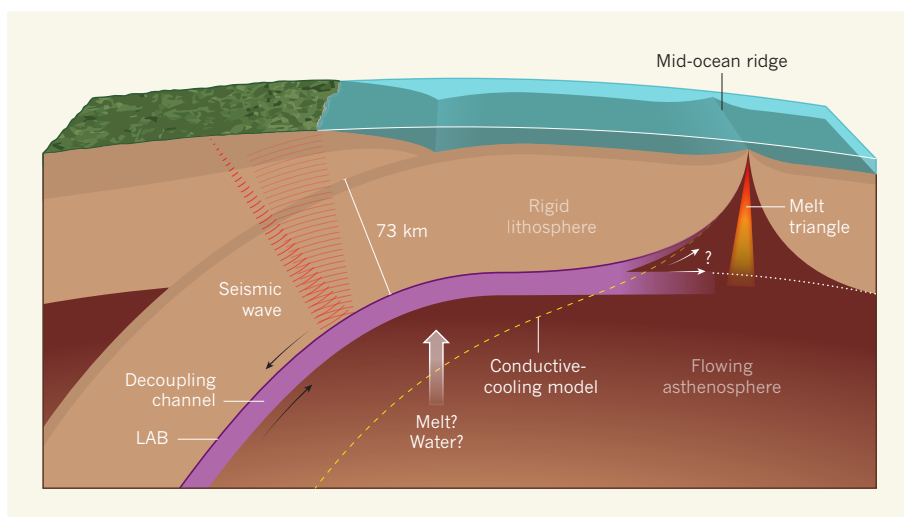

Figure 1 | A model of the subducting Pacific plate. Stern et al.[1] used seismic waves that travelled deep into the Earth and were reflected back to the surface to image the Pacific plate as it subducts beneath New Zealand. The seismic data reveal a plate that is 73 kilometres thick, significantly less than that predicted by conventional conductive-cooling models of the oceanic lithosphere (dashed line). The base of the plate, the lithosphere–asthenosphere boundary (LAB), is sharp (less than 1 km thick) and is underlain by a 10-km-thick water- or melt-rich channel that decouples the plate from the underlying flowing asthenosphere. The extent to which the channel continues beneath the ocean plate and/or varies in depth has yet to be determined, as indicated by the question mark. The arrows show the directions of motion in the different Earth layers. The plate may be relatively dehydrated, possibly as a result of melting beneath the mid-ocean ridge. The depth extent of the melt triangle would determine the thickness of the dehydrated layer (dotted line). Melt or water may travel to the channel from greater depths. The mechanism by which a 10-km-thick channel might form is unclear.

to melt. In this case, a steady supply of melt from greater depths in the mantle to the base of the plate would be required, given that the melt might travel up along the base of the plate.

The very existence of the channel itself is more of an enigma. How and why channelization would occur over a 10-km depth range is not known. Perhaps water availability from phase transformations[16] or melt ponding[15] occurs over a certain depth range. It could be specific to locations at which plate motions deviate from mantle flow, as is the case off both New Zealand and Costa Rica[17], where a similar channel was reported[18]. Overall, channels offer an explanation for some of the elusive nature of the LAB. Narrow channels would be nearly imperceptible in seismic imaging methods that rely on low-frequency waves, which might explain intermittent and discrepant LAB detection among methods[10]. For a full understanding of such channels, we need better constraints on where they exist.

However, global channel imaging may prove difficult. Anisotropy is probably important at the LAB, and may bias results if it is not properly considered. In addition, high seismic-wave frequencies are needed to distinguish fine-scale channels, although studies such as those of Stern and colleagues are not feasible at a global scale. Finally, what are the implications of these channels for the coupling of the plates to the underlying asthenosphere and the driving forces of plate tectonics? Tackling these questions will require incorporating

tight seismic constraints with laboratory experiments and geodynamical modelling. ∎

Catherine A. Rychert is at the National Oceanography Centre, Southampton, University of Southampton, Southampton SO14 3ZH, UK.
e-mail: c.rychert@soton.ac.uk

1. Stern, T. et al. Nature **518**, 85–88 (2015).
2. Rychert, C. A., Fischer, K. M. & Rondenay, S. Nature **436**, 542–545 (2005).
3. Hirth, G. & Kohlstedt, D. L. Earth Planet. Sci. Lett. **144**, 93–108 (1996).
4. Jackson, I., Faul, U. H., Fitz Gerald, J. D. & Morris, S. J. S. Mater. Sci. Eng. A **442**, 170–174 (2006).
5. McKenzie, D. J. Geophys. Res. **72**, 6261–6273 (1967).
6. Parsons, B. & Sclater, J. G. J. Geophys. Res. **82**, 803–827 (1977).
7. Rychert, C. A. & Shearer, P. M. J. Geophys. Res. **116**, B07307 (2011).
8. Burgos, G. et al. J. Geophys. Res. Solid Earth **119**, 1079–1093 (2014).
9. Beghein, C., Yuan, K. Q., Schmerr, N. & Xing, Z. Science **343**, 1237–1240 (2014).
10. Rychert, C. A., Schmerr, N. & Harmon, N. Geochem. Geophys. Geosyst. **13**, Q0AK10 (2012).
11. Kawakatsu, H. et al. Science **324**, 499–502 (2009).
12. Kodaira, S. et al. Nature Geosci. **7**, 371–375 (2014).
13. Karato, S.-I. Earth Planet. Sci. Lett. **321–322**, 95–103 (2012).
14. Gaherty, J. B., Kato, M. & Jordan, T. H. Phys. Earth Planet. Inter. **110**, 21–41 (1999).
15. Sakamaki, T. et al. Nature Geosci. **6**, 1041–1044 (2013).
16. Green, D. H., Hibberson, W. O., Kovács, I. & Rosenthal, A. Nature **467**, 448–451 (2010).
17. Becker, T. W., Conrad, C. P., Schaeffer, A. J. & Lebedev, S. Earth Planet. Sci. Lett. **401**, 236–250 (2014).
18. Naif, S., Key, K., Constable, S. & Evans, R. L. Nature **495**, 356–359 (2013).

CELL BIOLOGY

Organelles under light control

Optogenetic techniques enable light–activated control of protein–protein interactions in the cell. This approach has now been used to alter membrane dynamics and induce cellular reorganization. SEE LETTER P.111

FRANCK PEREZ

The internal organization of cells is finely tuned, and their intracellular dynamics ever-changing. This is particularly apparent in plant, animal and fungal cells, which contain specialized membrane-bound vesicles and organelles in which particular reactions occur. Studies show[1] that organelles must be correctly positioned to ensure proper cellular function, but it has so far been difficult or impossible to suddenly and reversibly alter the positions of vesicles and organelles in cells. In this issue, van Bergeijk et al.[2] (page 111) describe a technique that provides biologists with tools that can be thought of as light-responsive tweezers, enabling precise and rapid control of organelle positioning and movement in cells.

Organization of the cell's membranes depends on the cytoskeleton, and in particular on cytoskeletal microtubule and microfilament structures, which are involved in intracellular transport. Cells use these cytoskeletal networks, along with dedicated motor proteins such as kinesins, dyneins and myosins[3], to distribute and position different organelles in distinct subcellular regions[4]. For example, organelles called peroxysomes, which break down fatty-acid chains, are often distributed in the region around the nucleus. Early endosomes involved in cellular uptake of various cell-surface proteins are located at the cell periphery, whereas endosomes involved in molecule recycling are located at the cell centre. The mitochondria, which generate ATP molecules, are radially distributed along microtubules in the cytoplasm. Organelle

positioning can be even more extreme in highly polarized cells such as neurons and migrating cells, in which specialized functions must be sustained in specific parts of the cell, for example signalling at synaptic junctions or polarized secretion.

Optogenetic tools can regulate rapid and reversible interactions between selected protein domains in specific areas of a cell or tissue, in response to laser illumination[5–7]. In the current study, van Bergeijk et al. used these tools to induce the rapid recruitment of molecular motors to specific membranes. Several optogenetic systems currently exist, and the authors mostly used the 'LOV–ePDZb1' set-up[6]. This system is based on a modified light–oxygen–voltage (LOV) protein domain. The modified LOV domain changes conformation following illumination with blue light, blocking or unmasking an amino-acid motif through which the modified LOV interacts with another protein domain, ePDZb1.

The authors fused the modified LOV domain to proteins located on the membranes of specific organelles — PEX3 for peroxysomes, Rab11 for recycling endosomes and TOM20 for mitochondria. They fused the ePDZb1 domain to a motor protein, such as the kinesin KIF1A, which transports proteins to the cell periphery, or the BICD2 dynein-recruitment domain, which pulls proteins towards the cell centre. The myosin Vb motor-protein domain was also used to move membranes in neurons called dendrites. Thus, blue light induced indirect tethering of an organelle of interest to a desired motor protein, leading to movement of the organelle around the cell (Fig. 1). Conversely, van Bergeijk and colleagues transiently immobilized organelles and vesicles by fusing ePDZb1 either to the syntaphilin protein, which stably binds to microtubules, or to myosin-Vb, which stably binds the cytoskeleton in non-polarized cells.

Next, the researchers showed that this technique could alter organelle position almost instantaneously, changing the cell's organization in a matter of minutes. The reversibility of the system meant that it was possible to use intermittent cycles of illumination to study both organelle movement and the restoration of normal positioning. Laser illumination could be targeted to spots as small as 250 nanometres wide, as well as to much larger areas, meaning that perturbation of organelle positioning, or sudden immobilization of transport vesicles, could be achieved in specific cellular subdomains.

This system has many applications, as van Bergeijk and co-workers demonstrated. For example, they used local illumination to recruit BICD2 or KIF1A to Rab11-associated recycling endosomes in neuronal cells. This respectively decreased or increased the quantity of recycling endosomes in the growth cone (a structure located at the tip of growing
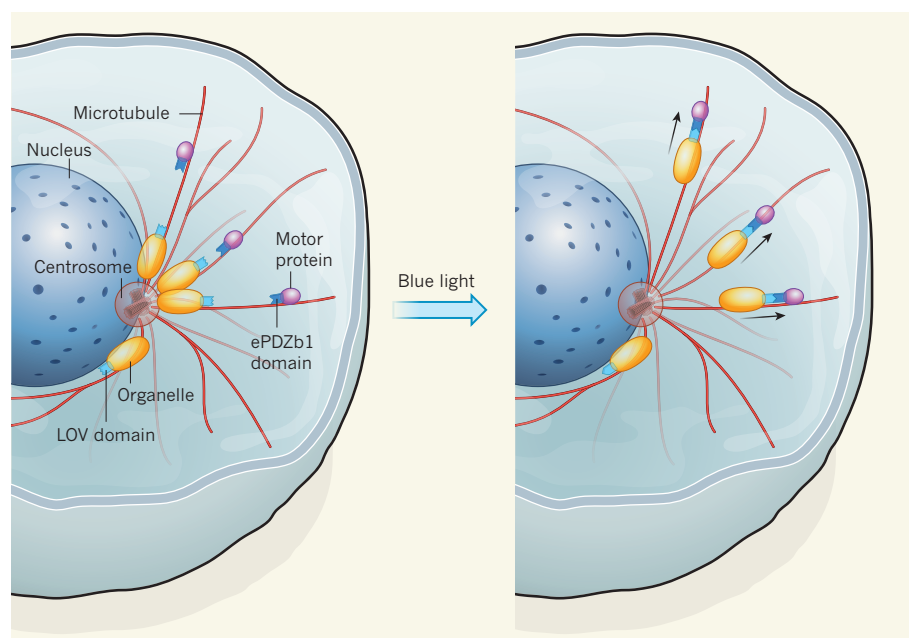

Figure 1 | A light touch. The position of subcellular structures such as organelles and vesicles is regulated by microtubules that are part of the cell's cytoskeleton. A system developed by van Bergeijk et al.[2] can be used to move organelles to different positions within the cell in response to light. The authors fused motor proteins to ePDZb1-protein domains, and organelle-associated proteins to a modified LOV-protein domain. The modified LOV domain changes conformation when exposed to blue light, unmasking an ePDZ-binding motif. Thus, following illumination, the modified LOV domain binds to ePDZb1. The organelle becomes tethered to the motor protein and its cellular position is altered by the motor protein's movement along microtubules — in the case shown here, from around the nucleus to the cell's periphery.

neuronal extensions, and in particular at the extremity of projections called axons). It has been shown[8,9] that Rab11 endosomes are involved in axon growth, but the authors' analysis goes further, showing that the presence of recycling endosomes in the growth cone is directly correlated with axon extension.

The authors also used their system to test the 'tug-of-war' model for positioning organelles in neuronal protrusions called dendritic spines. This model states that a balance between stable tethering and motor-driven forces is essential to regulate the positioning of organelles. Using their system, the researchers confirmed this model and defined the precise role of particular motor proteins and anchoring factors in polarized organelle trafficking. In summary, van Bergeijk and colleagues have designed a powerful tool that, with its high spatio-temporal resolution, is a spectacular example of the ability of optogenetics to alter processes in real time in chosen subcellular areas.

Technological revolutions have often provided the tools with which to analyse cellular processes from a different point of view. Examples include the advent of RNA interference, super-resolution fluorescent imaging, and electron microscopy and its subsequent improvements, all of which were instrumental in helping cell biologists to reimagine the cell. The next challenge is not only to improve existing tools, but also to develop additional approaches to asking new questions in a comprehensive and integrated manner.

Optogenetic strategies, including van Bergeijk and co-workers' technique, will have a major part to play here. For example, the quantitative spatio-temporal data that these techniques can generate will be of great use to fields such as systems biology and theoretical modelling. The study of cell biology at the tissue or whole-organism level will similarly benefit from such an approach, because it will be possible to suddenly change the positions and dynamics of specific organelles in particular cell types, and then monitor induced defects. Gene editing now allows us to create modified versions of key cellular regulatory factors[10]. Combining optogenetic development with gene editing will enable us to control cell organization precisely and to question its role in cellular function. A bright future awaits cell biology. ∎

Franck Perez *is at the CNRS UMR 144, Institut Curie, 75248 Paris 05, France.*
e-mail: franck.perez@curie.fr

1. Bornens, M. *Nature Rev. Mol. Cell Biol.* **9,** 874–886 (2008).
2. van Bergeijk, P., Adrian, M., Hoogenraad, C. C. & Kapitein, L. C. *Nature* **518,** 111–114 (2015).
3. Schliwa, M. & Woehlke, G. *Nature* **422,** 759–765 (2003).
4. de Forges, H., Bouissou, A. & Perez, F. *Int. J. Biochem. Cell Biol.* **44,** 266–274 (2012).
5. Gautier, A. *et al. Nature Chem. Biol.* **10,** 533–541 (2014).
6. Strickland, D. *et al. Nature Methods* **9,** 379–384 (2012).
7. Kennedy, M. J. *et al. Nature Methods* **7,** 973–975 (2010).
8. Bhuin, T. & Roy, J. K. *Cell Tissue Res.* **335,** 349–356 (2009).
9. Eva, R. *et al. J. Neurosci.* **30,** 11654–11669 (2010).
10. Hsu, P. D., Lander, E. S. & Zhang, F. *Cell* **157,** 1262–1278 (2014).

This article was published online on 7 January 2015.

DEVICE FABRICATION

Three-dimensional printed electronics

Can three-dimensional printing enable the mass customization of electronic devices? A study that exploits this method to create light-emitting diodes based on 'quantum dots' provides a step towards this goal.

JENNIFER A. LEWIS & BOK Y. AHN

The ability to rapidly print three-dimensional (3D) electronic devices would enable myriad applications, including displays, solid-state lighting, wearable electronics and biomedical devices with embedded circuitry. Writing in *Nano Letters*, Kong *et al.*[1] report an intriguing route to this goal by creating fully 3D-printed light-emitting diodes (LEDs) based on quantum dots. Quantum dots are semiconducting nanocrystals that exhibit tunable colour emission[2-4]. Using a 3D-printing method based on extruding multiple materials, the researchers patterned quantum-dot-based LEDs (QD-LEDs) on curved surfaces and integrated arrays of the diodes in 3D matrices.

3D printers transform the output files from computer-aided design tools into tangible objects using pattern-generating devices that move along multiple directions in space[4]. These devices can be light sources that harden resins or fuse powders, or nozzles that directly deposit materials. Since their introduction nearly three decades ago, 3D-printing methods have been used to build myriad objects, primarily prototypes, in a sequential, layer-by-layer fashion.

To create 3D objects of arbitrary form and specific function, a broad palette of materials and multi-material printing platforms are required. One promising approach is 3D extrusion printing[5], in which functional inks are deposited through fine cylindrical nozzles under an applied pressure at ambient conditions. Unlike 3D printers that use inkjet print heads, which are suitable only for inks with a narrow range of viscosities (about ten times that of pure water), extrusion-based printing enables materials of widely varying composition and viscosity to be patterned[6].

QD-LEDs are multilayer devices built around an active (light-emitting) layer composed of quantum dots[4]. Multiple QD-LED layer architectures have been explored with the aim of optimizing their external quantum efficiency, that is, the ratio of the number of photons emitted by the QD-LED relative to the number of electrons injected into the device

when an electric field is applied between the device's outer metallic electrodes (cathode and anode layers). In the typical embodiment of a QD-LED, the active layer is sandwiched between layers of electron- and hole-transporting materials, where holes are positively charged carriers. The applied electric field causes electrons and holes to move into the active layer, where they recombine to emit photons.

Solution-based processing routes have recently emerged for patterning QD-LEDs with the aim of lowering fabrication costs[7,8]. Central to this approach is the ability to maintain highly uniform layers between dissimilar materials. To create their QD-LED devices, Kong and colleagues sequentially printed several materials (Fig. 1). First, a conductive silver ring that surrounds a transparent anode layer followed by a hole-transport layer were printed, and annealed at 200 °C (silver ring) and 150 °C (other layers). Next, the active layer was formed by printing quantum dots suspended in a solvent mixture in a drop-wise manner. As solvent evaporation ensued, recirculating fluid flow suppressed quantum-dot migration to the drop edge, yielding relatively uniform, active layers[9]. Notably, each layer was patterned using immiscible solvents to minimize interlayer mixing. Finally, a cathode layer composed of liquid metal was printed[10] and the devices were packaged in a silicone sealant.

To highlight the flexibility of their approach, the authors printed QD-LEDs in multiple formats, including green and orange-red light emitters, 2 × 2 × 2 arrays embedded in a silicone matrix, and QD-LEDs on the surface of contact lenses and other substrates of interest.
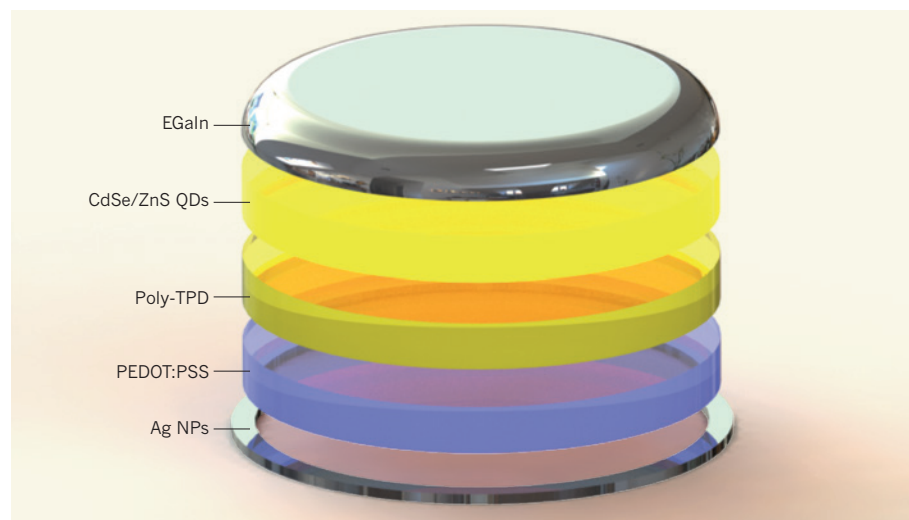
Figure 1 | Fully 3D-printed quantum-dot-based light-emitting diodes (QD-LEDs). The QD-LEDs reported by Kong and colleagues[1] consist of five layers: a conductive ring of silver nanoparticles (Ag NPs) that surrounds a transparent anode layer composed of poly(ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS); a hole-transport layer made of poly[N,N'-bis(4-butylphenyl)-N,N'-bis(phenyl) benzidine) (poly-TPD); a light-emitting layer composed of cadmium selenide/zinc sulfide quantum dots (CdSe/ZnS QDs); and a cathode layer composed of eutectic gallium indium (EGaIn). The diameter of the printed QD-LEDs is approximately 2 mm. (Figure adapted from ref. 1.)

The printed devices exhibit brightness, an essential metric of device performance, that is 10- to 100-fold below that of the best solution-processed QD-LEDs[3,8]. However, substantial improvements in device performance are likely to be possible by introducing an electron-transport layer (which was absent in the current architecture), such as one composed of zinc-oxide nanoparticles, and further optimizing the printing process.

The 3D-printing method used by the authors represents a simple, but sophisticated, approach for patterning functional materials. Demonstrated applications of this technique include printing electrodes that interconnect solar-cell and LED arrays[11], 3D antennas[12] and rechargeable microbatteries[13]. Although microbatteries rely on multi-material 3D printing of interdigitated cathode and anode layers, Kong and colleagues' study is much more impressive, because up to six, as opposed to two, different materials must be printed sequentially to create their devices.

One intriguing question that arises is whether fully 3D-printing electronic devices is the best approach for creating mass-customized electronics. Another viable strategy would be to combine 3D printing with automated pick-and-place machinery that places electronic components accurately and repeatably to generate objects with embedded circuitry and devices[11]. LEDs are commercially available that have lateral dimensions akin to those demonstrated by Kong *et al.*, and could be integrated into 3D-printed objects by this hybrid approach.

To vastly expand the capabilities of 3D printing, new functional inks and multi-nozzle print heads and printing platforms must be designed for rapidly and accurately patterning materials over a broad range of compositions and ink-flow behaviour. As these advances are realized, it may be possible to print customized 3D electronic devices in a highly scalable manner. We are becoming increasingly reliant on electronics in our daily lives, and so successful outcomes should be of great benefit to society. ■

Jennifer A. Lewis and **Bok Y. Ahn** *are in the School of Engineering and Applied Sciences and the Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, USA.*
e-mail: jalewis@seas.harvard.edu

1. Kong, Y. L. *et al. Nano Lett.* **14,** 7017–7023 (2014).
2. Colvin, V. L., Schlamp, M. C. & Alivisatos, A. P. *Nature* **370,** 354–357 (1994).
3. Coe, S., Woo, W.-K., Bawendi, M. & Bulović, V. *Nature* **420,** 800–803 (2002).
4. Shirasaki, Y., Supran, G. J., Bawendi, M. G. & Bulović, V. *Nature Photonics* **7,** 13–23 (2012).
5. Smay, J. E., Cesarano, J. III & Lewis, J. A. *Langmuir* **18,** 5429–5437 (2002).
6. Lewis, J. A. *Adv. Funct. Mater.* **16,** 2193–2204 (2006).
7. Wood, V. *et al. Adv. Mater.* **21,** 2151–2155 (2009).
8. Dai, X. *et al. Nature* **515,** 96–99 (2014).
9. Hu, H. & Larson, R. G. *J. Phys. Chem. B* **110,** 7090–7094 (2006).
10. Ladd, C., So, J.-H., Muth, J. & Dickey, M. D. *Adv. Mater.* **25,** 5081–5085 (2013).
11. Ahn, B. Y. *et al. Science* **323,** 1590–1593 (2009).
12. Adams, J. J. *et al. Adv. Mater.* **23,** 1335–1340 (2011).
13. Sun, K. *et al. Adv. Mater.* **25,** 4539–4543 (2013).

J.A.L. declares competing financial interests. See go.nature.com/benofm for details.

ECOLOGY

Deep and complex ways to survive bleaching

Mass coral bleaching events can drive reefs from being the domains of corals to becoming dominated by seaweed. But longitudinal data show that more than half of the reefs studied rebound to their former glory. SEE LETTER P.94

JOHN M. PANDOLFI

A constant battle for space is fought every minute of every day on the hard substrates that provide the foundation for living coral reefs. In one corner are reef corals and the photosynthetic dinoflagellate microalgae that live in symbiosis inside them; in the other are fleshy macroalgae, better known as seaweed. On healthy reefs, corals are the clear winners and dominate reef substrates (Fig. 1a). But regime shifts to macroalgae (Fig. 1b) often occur in response to local anthropogenic drivers such as overfishing of herbivores[1] or increased nutrients[2] from pollution and land-use changes — two conditions more favourable for seaweed than for corals. On page 94 of this issue, Graham *et al.*[3] provide the first

unequivocal evidence that regime shifts from corals to macroalgae also occur in response to coral bleaching, and they identify aspects of reef ecology that influence the likelihood of this occurring.

Coral bleaching occurs when the coral hosts expel their symbiotic dinoflagellates, which provide much of the vibrant coloration typical of coral reefs. Corals rely on the photosynthetic symbionts for their energy provision, and if bleached corals do not rapidly regain symbionts, they die. Mass bleaching events occur over broad spatial scales and affect a large component of the reef coral community. One such episode, in 1998, is often referred to as the largest mass bleaching event on record[4]; in the Seychelles, more than 90% of live coral cover was lost.
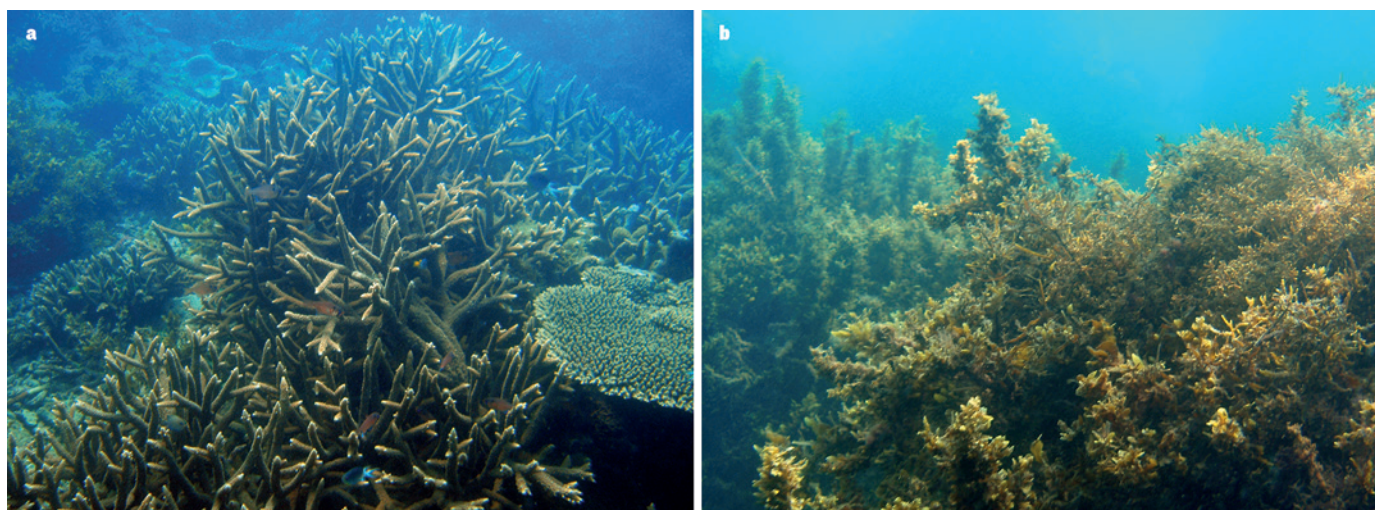
GEORGE ROFF

Figure 1 | Changing reefs. Graham *et al.*[3] show that mass coral bleaching events, such as the one that occurred in 1998, can drive reefs from being highly complex, coral-rich seascapes (**a**) to zones of dead coral dominated by macroalgae or seaweed (**b**). Such regime shifts had previously been known to occur only in response to local stressors such as overfishing or pollution.

Graham *et al.* tracked the response of coral and fish communities to this event across 21 inner Seychelles islands using a 17-year data set that started in 1994. They found that 9 of the 21 reefs underwent a regime shift to macroalgae, with the live coral cover decreasing from an average of 31% before the event to about 3% by 2011, and macroalgal cover increasing from 3% to 42% during the same period. Where these regime shifts occurred, the functional diversity of associated reef fishes shifted in concert with the changes in coral and macroalgal cover.

One of the key strengths of this study was its ability to test for predictors of ecosystem responses to the bleaching event. Graham and colleagues evaluated several potential factors: the three-dimensional structural complexity of the reef[5], water depth, abundance of juvenile corals, nutrient load, density of herbivorous fish and whether the reefs were part of 'no-take' marine reserves. The first three of these drivers turned out to be the most important. Indeed, combining structural complexity with water depth correctly predicted whether or not a regime shift would occur in 98% of cases — regime shifts occurred less frequently in more structurally complex and deeper-water habitats. These correlations bode well for our ability to predict the effects of future mass bleaching events, especially in tropical regions where conservation resources are limited, because these two variables can be quickly and easily measured on most reefs.

Coral reefs are often portrayed as one of the marine ecosystems that are most vulnerable to the threats of climate change, and global warming is commonly thought to be the principal underlying driver of mass bleaching events. Although Graham and colleagues' study is groundbreaking in its attribution of coral-to-algal regime shifts to a mass bleaching event, perhaps their most striking finding is that, in most cases (12 of 21 reefs), such regime shifts did not occur. The fact that more than half of the reefs fully recovered after the bleaching event is a promising outcome for the future of coral reefs. It is also consistent with studies showing that each mass bleaching leaves many sites unaffected, with almost complete recovery of corals from the 1998 event in many parts of the world[6], and that coral survivors of past bleaching events have a capacity to persist under subsequent bleaching events[7]. The findings also fit with experimental work suggesting that corals can quickly adapt to environmental change[8]. Put simply, many reef corals just might be capable of adapting fast enough to survive current rates of global environmental change[9,10].

A key challenge facing reef managers around the world is how to protect coral reefs from the 'big three' human threats: overfishing, pollution and climate change. A range of specific tools is available to tackle the first two of these, which are comparatively local stressors, but there is a paucity of appropriate climate-specific responses. Given the contribution of these local stressors to the global degradation of reefs, it is crucial that their management continues. However, Graham and colleagues' delineation of reef characteristics most closely associated with regime shifts caused by mass bleaching events means that we can now take concrete steps towards managing specifically for climate change as well. For example, the authors' findings suggest that structural complexity and water depth should be explicitly incorporated into the spatial design of marine reserves, with structurally complex and deep-water habitats targeted as high-value sites that will be more resistant to mass coral bleaching than shallower sites.

The authors' finding that the design of marine protected areas in the Seychelles had no bearing on the ability of reefs to rebound from the 1998 bleaching event is unsettling, and is a case in point of the need for new design approaches. But the Seychelles are not alone — many marine reserves only target areas that are important for sustaining fisheries. Perhaps we need to think about broadening the role of marine reserves to one that includes being a refuge from regime shifts, such that their success can be gauged not only by the number of fishes they contain, but also by the degree to which they protect explicit attributes of habitat diversity. To achieve this, Graham and colleagues' messages on how to manage reefs in the face of climate change will need to be placed in a global context, and further long-term studies from reefs in other regions will be needed if we are to fully understand the drivers of regime shifts on reefs. ∎

John M. Pandolfi *is in the School of Biological Sciences and the Australian Research Council Centre of Excellence for Coral Reef Studies, University of Queensland, Brisbane, Queensland 4072, Australia.*
e-mail: j.pandolfi@uq.edu.au

1. Hughes, T. P. *Science* **265,** 1547–1551 (1994).
2. Smith, J. E., Hunter, C. L. & Smith, C. M. *Oecologia* **163,** 497–507 (2010).
3. Graham, N. A. J., Jennings, S., MacNeil, M. A., Mouillot, D. & Wilson, S. K. *Nature* **518,** 94–97 (2015).
4. Wilkinson, C. in *Status of Coral Reefs of the World: 1998* (ed. Wilkinson, C.) 15–38 (Australian Inst. Mar. Sci., 1998).
5. Graham, N. A. J. & Nash, K. L. *Coral Reefs* **32,** 315–326 (2013).
6. Baker, A. C., Glynn, P. W. & Riegl, B. *Estuar. Coast. Shelf Sci.* **80,** 435–471 (2008).
7. Thompson, D. M. & van Woesik, R. *Proc. R. Soc. B* **276,** 2893–2901 (2009).
8. Palumbi, S. R., Barshis, D. J., Traylor-Knowles, N. & Bay, R. A. *Science* **344,** 895–898 (2014).
9. Pandolfi, J. M., Connolly, S. R., Marshall, D. J. & Cohen, A. L. *Science* **333,** 418–422 (2011).
10. Munday, P. L., Warner, R. R., Monro, K., Pandolfi, J. M. & Marshall, D. J. *Ecol. Lett.* **16,** 1488–1500 (2013).

This article was published online on 14 January 2015.

BIOCHEMISTRY

Elusive source of sulfur unravelled

The metabolic origin of the sulfur atom in the naturally occurring antibiotic lincomycin A has been obscure — until now. The biosynthetic steps involved reveal surprising roles for two sulfur–containing metabolites. SEE LETTER P.115

CHARLES E. MELANÇON III

Many vital biological molecules contain sulfur, the metabolic origins of which can be predicted fairly easily by those fluent in nature's biosynthetic language. But the sources of sulfur atoms are still difficult to predict for a few such molecules, including several with potentially useful biological activities, such as the anti-cancer agents calicheamicin[1] and leinamy-cin[2], and the antibacterial agent albomycin[3]. On page 115 of this issue, Zhao et al.[4] report the three key enzymatic steps that together install the sulfur atom of the antibiotic lincomycin A (Fig. 1) during its biosynthesis in the bacterium *Streptomyces lincolnensis*. Using a combination of techniques, they reveal a new role for mycothiol (MSH) — a bacterial sulfur-containing antioxidant — as a sulfur donor. They also report the unprecedented involvement of ergothioneine (EGT), another sulfur-containing bacterial metabolite, in an enzyme-catalysed process: the chemical activation of the carbon atom that will eventually bear the sulfur.

Deciphering the details of a complex biosynthetic pathway is, in some ways, like solving a jigsaw puzzle that has many more pieces than are needed to construct the correct picture. The molecular pieces (enzymes and their encoding genes, substrates and products)

must often be carefully examined to solve the puzzle successfully (determine the correct biosynthetic pathway). Some interesting mechanisms for incorporating sulfur into molecules have been revealed in the past few years, including those involved in the biosynthesis of thiamine[5] and the antibiotic BE-7585A (ref. 6).

The addition of whole-genome sequencing and sophisticated comparative genomics to the biosynthetic chemist's extensive repertoire of techniques has helped researchers to solve complex biosynthetic puzzles. Specifically, the method helps them to predict the relationships between enzyme amino-acid sequences and functions more accurately, and allows easier identification of genes involved in supplying biosynthetic precursors. This is particularly evident in genes found at chromosomal positions distant from those of related biosynthetic genes, which, in bacteria, are often tightly clustered. Zhao et al. use the full complement of tools to discover the elusive origin of the sulfur atom in lincomycin A.

As is often done in biosynthetic studies, the authors began by working backwards through the proposed pathway, attempting to reconstruct its molecular logic by isolating intermediates that accumulate when the activities of specific genes are disrupted. They proposed that *lmbE* — a gene present in the sequenced biosynthetic gene cluster for lincomycin A, and which encodes a homologue of an

amidase enzyme involved in an MSH-dependent detoxification process in certain bacteria — has a role in sulfur-atom incorporation.

Sure enough, when Zhao and colleagues disrupted the function of *lmbE*, they observed the accumulation of a lincomycin-like intermediate in which an MSH moiety was attached through its sulfur atom to the predicted sulfur-incorporation site. They also found that purified LmbE enzyme cleaves the amide bond of this intermediate, and that feeding either the isolated intermediate or the product of its cleavage by LmbE to an *S. lincolnensis* mutant in which the function of an MSH-regenerating gene, *mshA*, was disrupted led to restoration of lincomycin A production. The authors had identified *mshA* through genome sequencing and analysis as part of the study. These findings confirmed the function of LmbE and the intermediacy of its substrate and product in the biosynthetic pathway.

Next, the researchers examined the function of *lmbV*, another gene present in the lincomycin cluster. This gene is homologous to an enzyme that catalyses an MSH-dependent isomerization reaction. When Zhao and co-workers disrupted the function of *lmbV*, they observed the unexpected accumulation of another lincomycin-like intermediate, in which an EGT moiety is attached through its sulfur atom to the site occupied by MSH in the metabolite isolated from the *lmbE* mutant. The authors also identified the EGT-containing intermediate in their *mshA*-disruption mutant, which suggests that LmbV catalyses the replacement of the EGT moiety by MSH. They could not prove this directly, because they were not able to express and purify LmbV, but they confirmed their theory by showing that a homologue of LmbV — CcbV, an enzyme from a similar biosynthetic pathway[7] — catalyses the proposed reaction. They further confirmed the key role of EGT in the biosynthesis of lincomycin A by disrupting the function of *egtD*, a gene involved in the biosynthesis of EGT that they again found through genome mining.
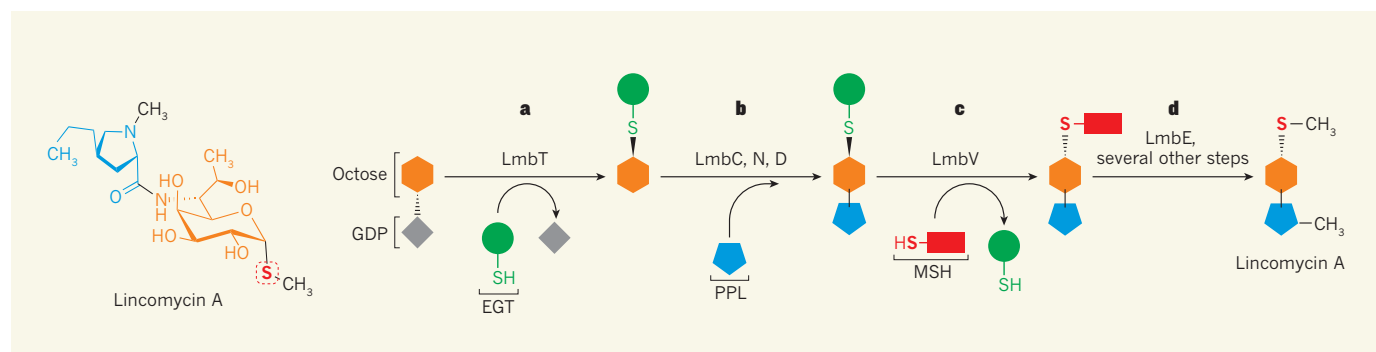

Figure 1 | Biosynthesis of lincomycin A. Zhao et al.[4] report the biosynthetic mechanism by which a sulfur atom is incorporated into the naturally occurring antibiotic lincomycin A. The two main components of the molecule are an octose sugar (orange) and a PPL unit (blue). **a**, In the biosynthetic pathway, the LmbT enzyme attaches ergothioneine (EGT, a bacterial metabolite) to the octose sugar by forming a new carbon–sulfur bond; a

nucleotide (GDP) attached to the sugar is lost in the process. **b**, The enzymes LmbC, N and D attach PPL to the sugar. **c**, LmbV then replaces EGT with mycothiol (MSH, a bacterial antioxidant), forming a new carbon–sulfur bond; the sulfur atom from MSH is the one that ends up in the antibiotic. **d**, Finally, a multi-step process beginning with a reaction catalysed by LmbE converts the MSH group into the methylmercapto group (–SCH₃) of lincomycin A.

50 Years Ago

It has been reliably demonstrated that rats can discriminate between the presence or absence of X-rays … The process by which X-rays elicit arousal and orienting reactions in mammals has not yet been determined. However, for simplicity we assume this mechanism operates via a 'radiation receptor'. Attempts to locate this hypothetical radiation receptor have yielded conflicting results … We used a narrow 3/16-in. X-ray beam as a signal or conditioned stimulus to warn the animal of a subsequent shock to the paws. The beam was most effective when it was directed at the olfactory region of the head … In an attempt to clarify this issue, we conducted a study of the effectiveness of X-ray as an arousing stimulus in rats the olfactory bulbs of which had been removed … The results indicate a distinct loss of sensitivity when the olfactory bulbs are removed.
From *Nature* 6 February 1965

100 Years Ago

In *Popular Astronomy* Prof. E. C. Pickering quotes some interesting letters from Profs. Backlund, of Pulkovo, and Schwarzchild, of Potsdam, with reference to astronomers and the war. None of the Pulkovo astronomers have been called to serve, but Prof. Backlund's son is in the Russian ranks, and of French astronomers, M. Croze, astrophysicist of the Paris Observatory, has been summoned, as well as the son of the director, M. Baillaud, who has six sons and sons-in-law in the war. On the German side, many young astronomers are in the field. Dr. Zurhellen and Dr. Kühl, who were with the eclipse expedition, have been interned in Russia; Dr. Münch, of Potsdam, is wounded and a prisoner in France.
From *Nature* 4 February 1915

Finally, Zhao *et al.* tested the function of another enzyme in the lincomycin pathway, LmbT, which they thought might attach EGT to a biosynthetic intermediate. LmbT is a homologue of glycosyltransferase enzymes, which attach sugars to other molecules. The researchers performed a series of gene-disruption and *in vitro* biochemical experiments, establishing that LmbT must act before installation of the 4-propyl-L-proline (PPL) moiety, which forms part of the structure of lincomycin A. In the process, they also proved that three more enzymes — LmbC (ref. 8), LmbN and LmbD — collectively incorporate PPL into the antibiotic.

Zhao and colleagues went on to isolate the suspected product of LmbT and to demonstrate the enzyme's function using *in vitro* assays. They discovered that LmbT catalyses the transfer of lincomycin A's sugar (for which the biosynthetic pathway has previously been reported[9]) to EGT, thus chemically activating the sugar in readiness for its reaction with MSH later in the pathway. Such a role is completely unprecedented: EGT was known to exist as a metabolite, but not as a substrate for an enzyme-catalysed reaction.

A particularly impressive aspect of this work is the authors' use of an intricate series of *in vivo* and *in vitro* experiments that relied on intermediates obtained from mutant cultures and from both enzymatic and chemical syntheses, guided by comparative gene analysis and genome mining. More generally, the study demonstrates that integration of primary metabolites (those that are essential for an organism's survival, such as MSH and EGT) and secondary metabolites (non-essential compounds, such as the products of the Lmb enzymes) is crucial for the biosynthesis of complex molecules. It also highlights the ingenious ways in which nature repurposes enzymes — in this case, using homologues of MSH-dependent detoxification enzymes for biosynthesis. And the establishment of functions for LmbE, LmbV and LmbT will no doubt help researchers to work out the functions of the enzymes' numerous homologues in the ever-growing roster of sequenced genomes. ∎

Charles E. Melançon III *is in the Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA.*
e-mail: cemelanc@unm.edu

1. Ahlert, J. *et al. Science* **297,** 1173–1176 (2002).
2. Tang, G. L, Cheng, Y. Q. & Shen, B. *Chem. Biol.* **11,** 33–45 (2004).
3. Zeng, Y. *et al. ACS Chem. Biol.* **7,** 1565–1575 (2012).
4. Zhao, Q., Wang, M., Xu, D., Zhang, Q & Liu, W. *Nature* **518,** 115–119 (2015).
5. Chatterjee, A. *et al. Nature* **478,** 542–546 (2011).
6. Sasaki, E. *et al. Nature* **510,** 427–430 (2012).
7. Čermák, L. *et al. Folia Microbiol.* **52,** 457–462 (2007).
8. Kadlčik, S. *et al. PLoS ONE* **8,** e84902 (2013).
9. Sasaki, E., Lin, C.-I., Lin, K.-Y. & Liu, H. *J. Am. Chem. Soc.* **134,** 17432–17435 (2012).

This article was published online on 14 January 2015.

PALAEOCLIMATE

Climate sensitivity in a warmer world

Comparison of climate records from the Pliocene and Pleistocene geological epochs of the past five million years suggests that positive climate feedbacks are not strengthened during warm climate intervals. SEE ARTICLE P.49

DAVID W. LEA

A major concern in projecting future climate change using models is that positive climate feedbacks might become enhanced in a warm climate, accelerating future warming in response to rising greenhouse-gas levels. Climate feedbacks are changes in atmospheric or surface properties induced by climate change that magnify or diminish the overall temperature response. Their aggregate strength is represented by the climate sensitivity, which is the ratio of observed warming to climate forcing, such as increasing atmospheric carbon dioxide levels. Warm intervals of Earth's recent geological past, which can be studied through climate proxies, provide a basis for testing the response of climate sensitivity to warming. On page 49 of this issue, Martínez-Botí *et al.*[1] use improved proxy atmospheric CO_2 data to compare climate-sensitivity determinations from the warm Pliocene epoch, 5.3 million to 2.6 million years (Myr) ago, to those from the cold, extensively glaciated Pleistocene epoch, 2.6 to 0.012 Myr ago. They find that climate sensitivity differs little between these vastly dissimilar times, once the influence of ice sheets is removed.

Why should climate sensitivity be stronger in a warm world? A warmer world is likely to have less snow and ice, thereby reducing their

Figure 1 | Then and now. Comparison of Strahcona Fjord, Ellesmere Island, in the high Canadian Arctic during the Pliocene epoch (illustration on the left) and today (photograph on the right). The changes illustrate the extreme polar amplification of warming during the Pliocene. Martínez-Botí *et al.*[1] use climate reconstructions from the Pliocene to determine whether positive climate feedbacks were stronger during warm intervals.

amplifying effect on climate change[2,3]. But how other feedbacks, such as water vapour and clouds, respond to warming is less certain. Simulations with climate models suggest that the positive feedback due to water vapour may strengthen in warmer climates[4], but uncertainties about how cloud feedbacks respond to warming confuse our understanding of the overall dependence of climate sensitivity on climate state.

Ancient climate records provide an alternative approach to assessing climate sensitivity, through the analysis of proxies, which reveal both the forcing (for example, atmospheric CO_2 levels or ice extent) and response (the temperature change). This approach offers the tremendous advantage of relying on natural equilibrium climate states rather than on synthetic ones simulated in models. Past climates were also influenced by various slow feedbacks such as ice sheets, vegetation and dust — factors typically not included in climate simulations. However, the method hinges on proxy reconstructions that have associated uncertainties, especially for marine-based atmospheric CO_2 reconstructions used in studies reaching beyond 0.8 Myr ago, the age of the oldest ice cores[5].

Palaeoclimate researchers have targeted the Pliocene epoch because it is the most recent time interval in which conditions were substantially warmer, about 2–3 °C warmer globally, than pre-industrial conditions[6]. Proxy reconstructions indicate that the Arctic climate during the Pliocene was much warmer than it is today, about 8–19 °C warmer, depending on location and season[7] (Fig. 1). But this extreme Arctic warmth seems to have coexisted, paradoxically, with atmospheric CO_2 levels that are similar to the present ones, implying an extreme amplification of positive climate feedbacks in the Pliocene[8].

Martínez-Botí and colleagues challenge this existing hypothesis using a well-validated technique to reconstruct Pliocene atmospheric CO_2 between 3.3 and 2.3 Myr ago at higher temporal resolution and with less variability than previous proxy reconstructions[5]. Their reconstruction clearly indicates for the first time that mid-Pliocene atmospheric CO_2 was up to 60% higher than pre-industrial values (450 parts per million (p.p.m.), compared with 280 p.p.m. before the Industrial Revolution and 400 p.p.m. today). The new record also indicates clear transitions in atmospheric CO_2 that are coherent with known climate events, including a drop in atmospheric CO_2 between 2.9 and 2.7 Myr ago that precedes global cooling and the onset of Northern Hemisphere glaciation 2.6 Myr ago — remarkable findings in themselves.

The researchers go a step further, applying their results to climate sensitivity for the warm Pliocene state by developing averaged reconstructions of land and ocean temperature and comparing them directly to their atmospheric CO_2 reconstruction. The slope relating the forcing (atmospheric CO_2) and response (temperature) at each time slice yields a tight constraint on climate sensitivity that is specific to the Pliocene. The authors find Pliocene climate sensitivity to be half as strong as that found for the cold Pleistocene. A repeat of the analysis after removing the forcing associated with glacial–interglacial changes in ice sheets reveals that Pliocene and Pleistocene climate sensitivities to atmospheric CO_2 changes alone were essentially the same.

In a broader context, these results also relate to attempts to use the instrumental temperature record to narrow the range of equilibrium climate sensitivity, which is the equilibrium temperature change caused by a doubling of atmospheric CO_2 allowing for 'fast' feedback processes only. Some studies have argued[9] that the slightly weaker rate of global warming since 2001 reduces the lower boundary of equilibrium climate sensitivity to well below 2 °C. Although Martínez-Botí and colleagues' derived Earth-system sensitivity[10] includes slow feedbacks, which complicates direct comparison to results from climate models, their results are likely to translate[10] to an equilibrium climate sensitivity of between 2 and 3 °C, well within the generally accepted range.

Despite the significant advance Martínez-Botí and co-workers' study represents, several challenges remain. First, given the wide range of proxy atmospheric CO_2 data for the Pliocene[5], it will be essential to validate the new results and assess why earlier reconstructions and methodologies differ from this one. Second, the extreme Pliocene warming in the terrestrial Arctic (Fig. 1) still requires enhanced polar climate feedbacks that remain unexplained[7]. Finally, for climate modellers, there remains the substantial challenge of reconciling emerging palaeoclimate-based sensitivity results with simulations of both past and future climate states. ∎

David W. Lea *is in the Department of Earth Science, University of California, Santa Barbara, California 93106-9630, USA.*
e-mail: lea@geol.ucsb.edu

1. Martínez-Botí, M. A. *et al. Nature* **518,** 49–54 (2015).
2. Colman, R. & McAvaney, B. *Geophys. Res. Lett.* **36,** L01702 (2009).
3. Kutzbach, J. E., He, F., Vavrus, S. J. & Ruddiman, W. F. *Geophys. Res. Lett.* **40,** 3721–3726 (2013).
4. Meraner, K., Mauritsen, T. & Voigt, A. *Geophys. Res. Lett.* **40,** 5944–5948 (2013).
5. Zhang, Y. G., Pagani, M., Liu, Z., Bohaty, S. M. & DeConto, R. *Phil. Trans. R. Soc. A* **371,** 20130096 (2013).
6. Dowsett, H. *et al. Sci. Rep.* **3,** 2013 (2013).
7. Brigham-Grette, J. *et al. Science* **340,** 1421–1427 (2013).
8. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. *Nature Geosci.* **3,** 27–30 (2010).
9. Otto, A. *et al. Nature Geosci.* **6,** 415–416 (2013).
10. Lunt, D. J. *et al. Nature Geosci.* **3,** 60–64 (2010).

ARTICLE

Plio–Pleistocene climate sensitivity evaluated using high-resolution CO$_2$ records

M. A. Martínez-Botí[1]*, G. L. Foster[1]*, T. B. Chalk[1], E. J. Rohling[1,2], P. F. Sexton[3], D. J. Lunt[4,5], R. D. Pancost[5,6], M. P. S. Badger[5,6] & D. N. Schmidt[5,7]

Theory and climate modelling suggest that the sensitivity of Earth's climate to changes in radiative forcing could depend on the background climate. However, palaeoclimate data have thus far been insufficient to provide a conclusive test of this prediction. Here we present atmospheric carbon dioxide (CO$_2$) reconstructions based on multi-site boron-isotope records from the late Pliocene epoch (3.3 to 2.3 million years ago). We find that Earth's climate sensitivity to CO$_2$-based radiative forcing (Earth system sensitivity) was half as strong during the warm Pliocene as during the cold late Pleistocene epoch (0.8 to 0.01 million years ago). We attribute this difference to the radiative impacts of continental ice-volume changes (the ice-albedo feedback) during the late Pleistocene, because equilibrium climate sensitivity is identical for the two intervals when we account for such impacts using sea-level reconstructions. We conclude that, on a global scale, no unexpected climate feedbacks operated during the warm Pliocene, and that predictions of equilibrium climate sensitivity (excluding long-term ice-albedo feedbacks) for our Pliocene-like future (with CO$_2$ levels up to maximum Pliocene levels of 450 parts per million) are well described by the currently accepted range of an increase of 1.5 K to 4.5 K per doubling of CO$_2$.

Since the start of the industrial revolution, the concentration of atmospheric CO$_2$ (and other greenhouse gases) has increased dramatically (from ~280 to ~400 parts per million)[1]. It has been known for over 100 years that changes in greenhouse gas concentration will cause the surface temperature of Earth to vary[2]. A wide range of observations reveals that the sensitivity of Earth's surface temperature to radiative forcing amounts to ~3 K warming per doubling of atmospheric CO$_2$ concentration (with a 66% confidence range of 1.5–4.5 K; see refs 1 and 3), caused by direct radiative forcing by CO$_2$ plus the action of a number of fast-acting positive feedback mechanisms, mainly related to atmospheric water vapour content and sea-ice and cloud albedo. Uncertainty in the magnitude of these feedbacks confounds our ability to determine the exact equilibrium climate sensitivity (ECS; the equilibrium global temperature change for a doubling of CO$_2$ on timescales of about a century, after all 'fast' feedbacks have had time to operate; see ref. 3 for more detail). Although the likely range of values for ECS is 1.5–4.5 K per CO$_2$ doubling, there is a small but finite possibility that climate sensitivity may exceed 5 K (see ref. 1). Understanding the likely value of ECS clearly has important implications for the magnitude, eventual impact and potential mitigation of future climate change.

Any long-range forecast of global temperature (that is, beyond the next 100 years) must also consider the possibility that ECS could depend on the background state of the climate[4,5]. That is, in a warmer world, some feedbacks that determine ECS could become more efficient and/or new feedbacks could become active to give additional warmth for a given change in radiative forcing (such as those relating to methane cycling[6], atmospheric water vapour concentrations[5,7,8], in addition to changes in the relative opacity of CO$_2$ to long-wave radiation[5,9]). One approach to identify whether ECS depends on the climate background state is to

reconstruct ECS during periods in the geological past when Earth was warmer than today.

The Pliocene (2.6–5.3 million years (Myr) ago) is one such time, with the warmest intervals between 3.0 Myr and 3.3 Myr ago about 3 K globally warmer than pre-industrial times[10,11]; the mean sea level was 12–32 m above the present level[12,13]. Although most of this warmth is commonly ascribed to increased atmospheric CO$_2$ levels[14], it has been suggested that simple comparisons of the observed temperature change in the geological record with the climate forcing from CO$_2$ alone are unable to constrain ECS[10]. Instead, a parameter termed the Earth system sensitivity (ESS) is defined, as the change in global temperature for a doubling of CO$_2$ once both fast and slow feedbacks have acted and the whole Earth system has reached equilibrium. (In contrast, ECS excludes the slow feedbacks; for a discussion of fast versus slow feedbacks, see ref. 3.) The most important slow feedbacks are those related to ice-albedo and vegetation-albedo changes. Because of these slow feedbacks, Pliocene ESS is thought to have been ~50% higher than ECS[10,15], with some existing geological data suggesting a Pliocene ESS range of 7–10 K per CO$_2$ doubling[16], which greatly exceeds a modern ESS estimate of ~4 K per CO$_2$ doubling[10]. If ECS was similarly enhanced, then that would imply that either extra positive fast feedbacks operated, or that existing positive fast feedbacks were more efficient, thus increasing the temperature response for a given level of CO$_2$ forcing.

Understanding past climate sensitivity depends critically on the accuracy of the CO$_2$ data used. Despite a tendency towards increased agreement between different CO$_2$ proxies[17], individual estimates of the partial pressure of CO$_2$ (p_{CO_2}) for the Pliocene still range from ~190 μatm to ~440 μatm (Fig. 1a, b) and there is little coherence in the trends described by the various techniques (Fig. 1a, b). This hinders any effort to accurately

[1]Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, Southampton, SO14 3ZH, UK. [2]Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. [3]Centre for Earth, Planetary, Space and Astronomical Research, The Open University, Milton Keynes, MK7 6AA, UK. [4]School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK. [5]The Cabot Institute, University of Bristol, Bristol BS8 1UJ, UK. [6]Organic Geochemistry Unit, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. [7]School of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol, BS8 1RJ, UK.
*These authors contributed equally to this work.

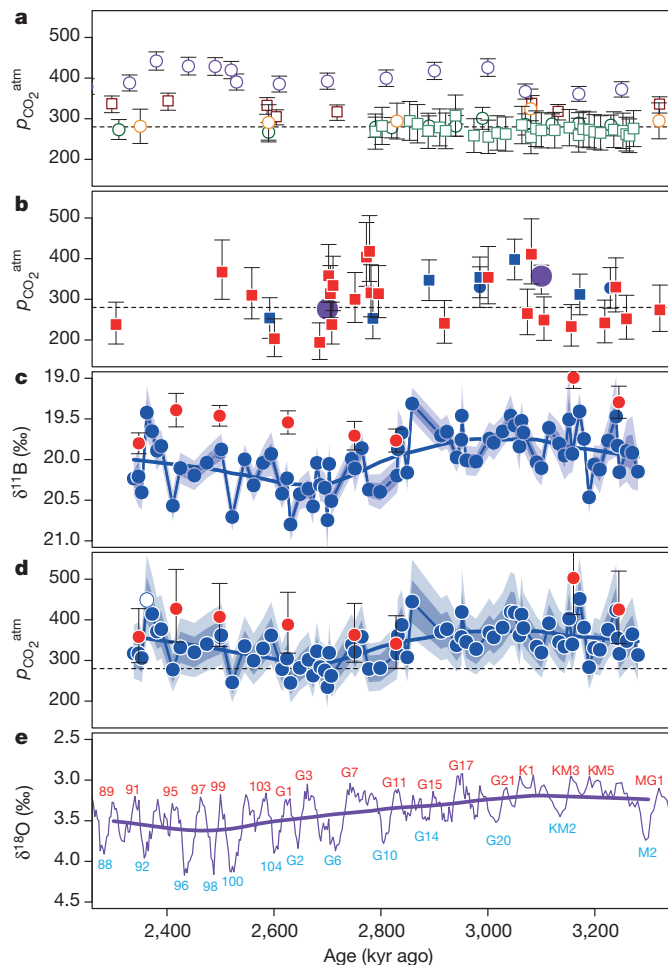**Figure 1 | Records of late Pliocene/early Pleistocene $p_{CO_2}^{atm}$. a**, $p_{CO_2}^{atm}$ based on $\delta^{13}C$ of sedimentary alkenones (dark green circles (ODP 999)[25]; aquamarine squares (ODP 999)[26]; dark orange (ODP 1208)[16], purple circles (ODP 806)[16]; and dark red squares (ODP 925)[49]). Error bars are uncertainty in $p_{CO_2}^{atm}$ at the 95% level of confidence. **b**, $\delta^{11}B$ of planktic foraminifera from ODP 999 (blue circles for *Globigerinoides sacculifer*[25] and blue squares for *G. ruber*[25]; red squares for *G. sacculifer*[23]) and stomatal density of fossil leaves (purple filled circle)[50]. Error bars are uncertainty in $p_{CO_2}^{atm}$ at the 95% level of confidence. **c**, New boron isotope data (this study) from ODP 999 (blue circles) and ODP 662 (red circles). Error bands for ODP 999 denote 1 standard deviation (sd) (dark blue) and 2 sd (light blue) analytical uncertainty; error bars for ODP 662 show 2 sd analytical uncertainty. **d**, Atmospheric p_{CO_2} determined from data shown in **c** for ODP 999 (blue circles) and ODP 662 (red circles). Error band encompasses 68% (dark blue) and 95% (light blue) of 10,000 Monte Carlo simulations of $p_{CO_2}^{atm}$ using the data in **c** and a full propagation of all the key uncertainties (see Methods). For ODP 662 error bars encompass 95% of 10,000 simulations. Dotted lines show the modelled threshold of Northern Hemisphere glaciation (280 µatm)[33]. **e**, Benthic $\delta^{18}O$ stack[22], with prominent marine isotope stages labelled (blue for glacial, red for interglacial stages). Thick lines on several panels are non-parametric smoothers through the data. The blue open circle in **d** highlights the data point that is identified as an outlier in Fig. 2 and not used in subsequent regressions.

constrain Pliocene ECS or ESS. To better determine Pliocene CO₂ levels, we generated a new record, based on the boron isotopic composition ($\delta^{11}B$) of the surface mixed-layer dwelling planktic foraminiferal species *Globigerinoides ruber* from Ocean Drilling Program (ODP) Site 999 (Caribbean Sea, 12° 44.64′ N, 78° 44.36′ W, 2,838 m water depth; Extended Data Fig. 1) at a temporal resolution (one sample about every 13,000 years (13 kyr); Fig. 1c) that is more than three times higher than previous $\delta^{11}B$ records (one sample every 50 kyr; Fig. 1b). The $\delta^{11}B$ of *G. ruber* is a well-constrained function of pH (ref. 18) and seawater pH is well correlated with $[CO_2]_{aq}$ (the aqueous concentration of CO_2),

because both are a function of the ratio of alkalinity to total dissolved carbon in seawater. In the absence of major changes in surface hydrography, $[CO_2]_{aq}$ is largely a function of atmospheric CO_2 levels, and $\delta^{11}B$-derived CO_2 has been demonstrated to be an accurate recorder of atmospheric CO_2 (Extended Data Fig. 2a)[18–20]. Today, the surface water at ODP Site 999 is close to equilibrium with the atmosphere with respect to CO_2 (expressed here as $\Delta p_{CO_2} = p_{CO_2}^{sw} - p_{CO_2}^{atm} = +21$ µatm; Extended Data Fig. 1)[18,21] and has remained so for at least the past 130 kyr (Extended Data Fig. 2)[18]. ODP Site 999 also benefits from a detailed astronomically calibrated age model[22] and high abundance of well-preserved planktic foraminifera throughout the past 4 Myr (refs 23, 24). During our study interval it is also unlikely to have been influenced by long-term oceanographic changes such as the emergence of the Panama isthmus ~3.5 Myr ago (see detailed discussion in ref. 23). To increase confidence that atmospheric CO_2 changes are driving our pH (and hence our $p_{CO_2}^{sw}$) record for ODP Site 999 and that the air–sea CO_2 disequilibrium remained similar to modern values, we also present lower-resolution $\delta^{11}B$ data from *G. ruber* from ODP Site 662 (equatorial Atlantic, Fig. 1c; 1° 23.41′ S, 11° 44.35′°W, 3,821 m water depth; Extended Data Fig. 1), where current mean annual Δp_{CO_2} is +29 µatm with a seasonal maximum of +41 µatm (ref. 21). Analytical methodology and information detailing precisely how $p_{CO_2}^{sw}$ is calculated, with full propagation of uncertainties, can be found in the Methods (with full $\delta^{11}B$ and p_{CO_2} values listed in Supplementary Tables 1 and 2).

A new record of Pliocene p_{CO_2} change

Where our data for both sites overlap in time, reconstructed $p_{CO_2}^{atm}$ values 2.3–3.3 Myr ago agree within uncertainty (Fig. 1d; Extended Data Fig. 3), and are consistent with most independent records (see Fig. 1a, b; Extended Data Fig. 2b, c), confirming that the variations we observe are predominantly driven by changes in atmospheric CO_2 concentrations. However, the enhanced resolution of our $\delta^{11}B$–$p_{CO_2}^{atm}$ record (Fig. 1d) also reveals a hitherto undocumented[16,23,25,26] level of structure in the CO_2 variability during the one-million-year period investigated, including a transition centred on 2.8 Myr ago, spanning ~200 kyr, during which average $p_{CO_2}^{atm}$ undergoes a decrease of ~65 µatm (Fig. 1d).

Detailed atmospheric CO_2 measurements from ice cores show orbital-scale (~100 kyr) oscillations in $p_{CO_2}^{atm}$ with a peak-to-trough variation of ~80–100 µatm through the late Pleistocene (90% of the p_{CO_2} values lie between +36 µatm and –41 µatm of the long-term mean; Extended Data Figs 2 and 4)[27–29]. Once the long-term trend is removed from our Plio-Pleistocene data (thick blue line in Fig. 1d), and we have taken into account our larger analytical uncertainty (see Methods), we observe orbital-scale variations in our $\delta^{11}B$–$p_{CO_2}^{atm}$ record of only slightly smaller amplitude than the ice-core $p_{CO_2}^{atm}$ record (0–0.8 Myr) and for the last 2 Myr in other $\delta^{11}B$-based records[19,20,30] (Extended Data Fig. 4 and Methods), which is in clear contrast with the benthic $\delta^{18}O$, which shows increasing variability over the last 3 Myr (Fig. 1e and Extended Data Fig. 4).

Given the different amplitudes of climate variability, the observed similarity between Pliocene and late Pleistocene $p_{CO_2}^{atm}$ variability seems counter-intuitive given the notion that CO_2 is a key factor in amplifying glacial–interglacial climate change[27–29,32]. This is illustrated by a well-defined nonlinear relationship in a cross plot between deep-sea benthic $\delta^{18}O$ and $ln(CO_2/C_o)$ (where C_o is the pre-industrial CO_2 level of 278 µatm), which accounts for the logarithmic nature of the climate forcing by CO_2 (Fig. 2b). Note also the clear overlap between Pleistocene (0–2.2 Myr) ice-core CO_2 measurements and $\delta^{11}B$-based CO_2 reconstructions in this plot (Fig. 2b; Extended Data Fig. 2). A similar relationship is also evident in raw $\delta^{11}B$-space (Fig. 2a). Below an inflection at about $p_{CO_2}^{atm} = 275 \pm 15$ µatm (equating to $ln(CO_2/C_o) \approx 0$), benthic $\delta^{18}O$ shows a steeper relationship with CO_2-based forcing than it does above this value (Fig. 2). This probably reflects some combination of: (1) growth of larger Northern Hemisphere ice sheets at $p_{CO_2}^{atm}$ below 275 ± 15 µatm (ref. 33), increasing radiative ice-albedo feedback and amplifying climate forcing by CO_2 change; (2) an increase in oxygen isotope fractionation in precipitation with increasing size of the
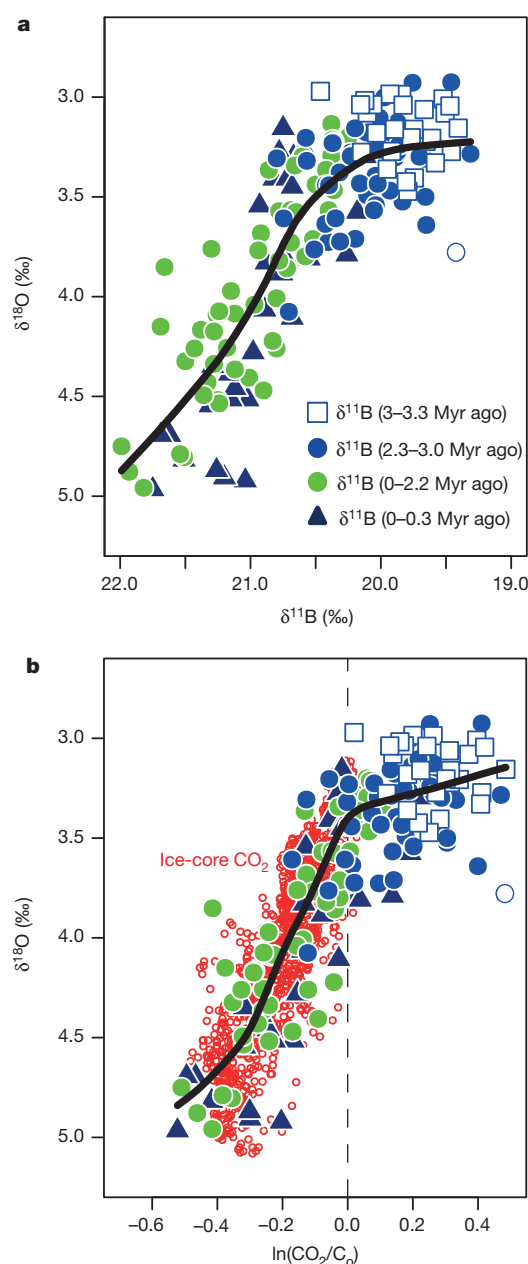
Figure 2 | Relationship between $\delta^{11}B$, climate forcing from CO_2 and $\delta^{18}O$. a, b, $\delta^{11}B$ versus $\delta^{18}O$ (**a**) and $\ln(CO_2/C_o)$ versus $\delta^{18}O$ (**b**) for data from the last 3 Myr. Boron data in **a** are from this study (blue open and closed circles) and published studies (green circles[30]; blue triangles[20]). Ice-core CO_2 data are shown as open red circles[27–29]. The vertical dashed line is at a CO_2 of 278 μatm. The data point removed from subsequent regression analysis is shown as open blue circles in **a** and **b**. Note that for clarity the $\delta^{11}B$–p_{CO_2} data from ref. 23 are not plotted. The black line is a non-parametric regression through all the data shown. The $\delta^{11}B$ data from ref. 30 have been corrected for laboratory and inter-species differences through a comparison between core-top $\delta^{11}B$ values.

ice sheets, which leads to a proportionally greater ^{18}O enrichment in seawater[34]; and (3) potentially stronger deep-sea cooling at low $p_{CO_2}^{\text{atm}}$ due to the high-latitude-focused influences of the ice-albedo feedback process. These findings highlight the profound impacts of Northern Hemisphere ice-sheet growth on climate variability in the Pleistocene[31,32], relative to the Pliocene (Fig. 2b).

Our data show that the ~275 ± 15 μatm threshold was first crossed ~2.8 Myr ago during Marine Isotope Stage (MIS) G10 (Fig. 1d, horizontal dashed line) and—more persistently—during subsequent MISs G6 (2.72 Myr ago), G2 (2.65 Myr ago), and 100 (2.52 Myr ago), when values

as low as 233^{+63}_{-53} μatm (95% confidence) were reached and when intervening interglacial values also seem to have been suppressed (Fig. 1c, d). These isotope stages are notable in that they are associated with an increase in the amplitude of glacial–interglacial sea-level oscillations (Extended Data Fig. 5b)[12,13,35] and coincide with the timing of the first substantial continental glaciations of Europe, North America and the Canadian Cordillera, as reconstructed by ice-rafted debris and observations of relic continental glacial deposits[36–38]. Hence, our high-resolution $p_{CO_2}^{\text{atm}}$ record robustly confirms previous hypotheses[16,23,25,39] (based on low-resolution CO_2 data) that the first substantial stages of glaciation in the Northern Hemisphere, as well as a recently recognized deep-sea cooling during the late Pliocene/early Pleistocene[13], coincided with a significant decline in mean atmospheric $p_{CO_2}^{\text{atm}}$ 2.7–2.9 Myr ago of ~40–90 μatm (the mean value of $p_{CO_2}^{\text{atm}}$ 3.0–3.2 Myr ago minus its mean 2.4–2.7 Myr ago = 66 ± 26 μatm; two-tailed $P < 0.001$, $n = 40$).

Efficiency of climate feedbacks

The high fidelity of the boron isotope pH/$p_{CO_2}^{\text{atm}}$ proxy (Extended Data Fig. 2), coupled with the high resolution of our $p_{CO_2}^{\text{atm}}$ record, offers an opportunity to examine the sensitivity of Earth's climate system to forcing by CO_2 during a period when Earth's climate was, on average, warmer than today[40]. For this exercise, global temperature estimates are also needed. We consider two approaches for this. The first is an estimate of global mean annual surface air temperature change (ΔMAT) over the last 3.5 Myr, from a scaling of the Northern Hemisphere climate required to drive an ice-sheet model to produce deep-ocean temperature and ice-volume changes consistent with benthic $\delta^{18}O$ data (Fig. 3a, b)[35]. This approach produces a continuous record of global temperature that agrees well with independent constraints for discrete time intervals (see ref. 35).

We supplement ΔMAT with a record from a second approach, which is independent from benthic $\delta^{18}O$ values. For this, we generated a sea surface temperature stack (SST_{st}) from 0 to 3.5 Myr ago (Fig. 3c, d), comprising ten high-resolution (average ~4 kyr) SST records based on alkenone unsaturation ratios (expressed as the $U_{37}^{K'}$ palaeotemperature index), from latitudes between 41 °S and 57 °N. The selected sites (see Extended Data Fig. 1b) all offer near-continuous temporal coverage of the last 3.5 Myr (see Methods). Our SST stack agrees well with independent, higher-density compilations of global SST change[32,41] (blue line in Fig. 3c), indicating that SST_{st} offers a reliable approximation of global SST change (see Methods for more details). Moreover, our SST stack allows us to directly compare the major SST changes, within the same archives, between the Plio-Pleistocene and late Pleistocene.

When comparing temperature records from the two approaches considered, it must be emphasized that ΔMAT reflects the global mean annual surface air temperature change, while the SST stack approximates the global mean sea surface temperature change. Hence, their amplitudes of variability will be different, mainly because SST_{st} does not include temperature changes over land. Approximately, $\Delta SST = \Delta MAT \times 0.66$ (refs 32 and 42), but a direct conversion is not needed here, because we merely aim to contrast Pliocene climate behaviour with that for the Pleistocene, within the same data types.

To determine the sensitivity of global SST and ΔMAT to CO_2 forcing in the Pliocene and Pleistocene, we use a time series of forcing calculated from our new and existing CO_2 records (Fig. 3e–h), and regress these against both ΔMAT and the SST stack (Fig. 3a to d; Supplementary Tables 1–3). The regression slopes then describe the average temperature change ΔT (in units of K) per watt per square metre of forcing (ΔF) for each time interval. These gradients therefore approximate the commonly used sensitivity parameter $S = \Delta T/\Delta F$ (in units of $K W^{-1} m^2$) for describing global temperature change for a given forcing[3]. In this scheme, a doubling of atmospheric CO_2 is equivalent to a forcing of $3.7 W m^{-2}$, so that for the 66% confidence interval of modern climate sensitivity quoted by ref. 1, the present-day equilibrium value of S (which we denote S^a, where superscript 'a' is for 'actuo', after ref. 3) is $1.5/3.7$ to $4.5/3.7 = 0.4–1.2 K W^{-1} m^2$. However, using palaeoclimate
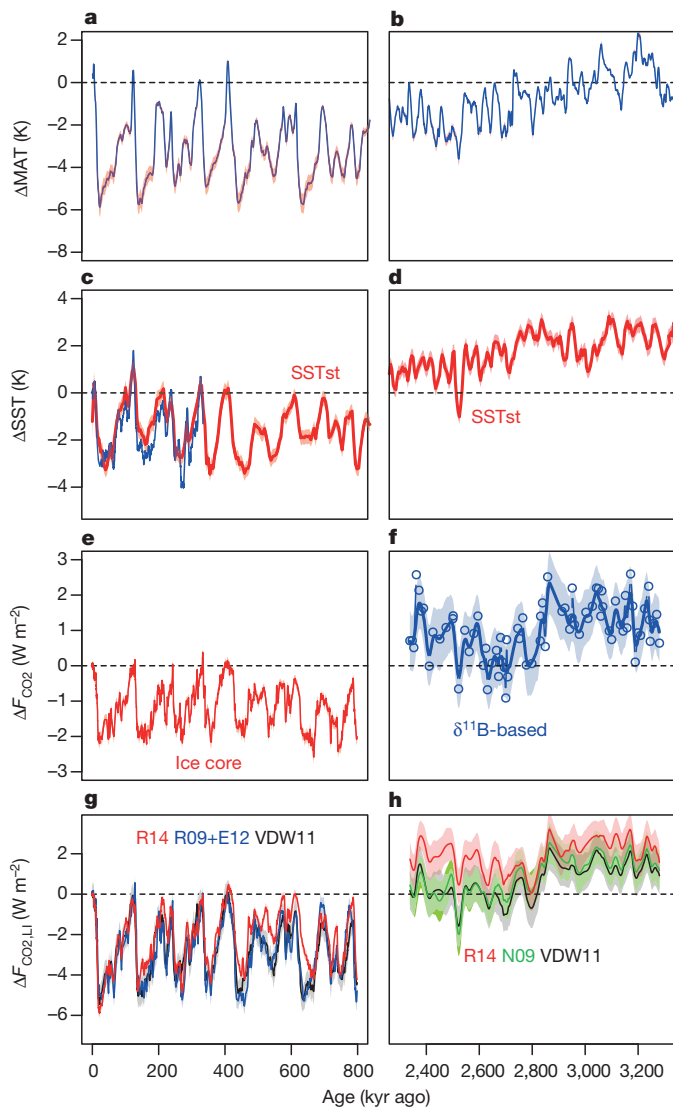
Figure 3 | Pleistocene and late Pliocene time series. a, b, ΔMAT[35]. **c, d,** ΔSST.
Data from this study are shown in red and from a stack of a more
comprehensive compilation[32] in blue. Uncertainty envelopes at 95% confidence
for both temperature records are shown in red. **e,** ΔF_{CO2} for the Pleistocene
from ice-core data[27–29]. **f,** ΔF_{CO2} for the late Pliocene, calculated using the
CO_2 data from this study. **g,** $\Delta F_{CO2,LI}$ calculated using data in **e** and published
sea-level records (R14 from ref. 13, VDW11 from ref. 35 and R09+E12
from ref. 44 for 0–520 kyr and ref. 45 for 520–800 kyr). **h,** $\Delta F_{CO2,LI}$ for the
late Pliocene calculated using the CO_2 data from this study and published sea-
level records (N09 from ref. 46 recalculated by ref. 12, R14 from ref. 13, VDW11
from ref. 35). Error bands in **e** to **h** represent the uncertainty in smoothed
CO_2 record and sea level (at 95% confidence), propagated using a Monte Carlo
approach ($n = 1,000$) for each reconstruction.

data it is not possible to determine the direct equivalent of S^a; instead,
such studies constrain a 'past' parameter (S^P), which includes the com-
bined action of both fast and slow feedbacks[3]. Note that ESS (in units
of K) $= S^P \times 3.7$. Explicitly accounting for slow feedback processes in
determinations of S^P can make it approximate S^a (ref. 3). Following ref. 3,
an S^P estimate after accounting for carbon-cycle feedback is indicated
by S_{CO2}, and one accounting for both carbon-cycle feedbacks and feed-
backs between land-ice and albedo feedbacks is $S_{CO2,LI}$, where the latter
gives a useful approximation of S^a. We follow this approach, using $S^P =$
$\Delta MAT/\Delta F$ and $S^{P,SST} = \Delta SST/\Delta F$ (both in units of K W^{-1} m^2). Note
that our determinations of the sensitivity parameter are based on our
entire reconstructed time series, rather than on a simple comparison
between a limited Pliocene average and the modern average, as was

done in previous studies[3,16]. Since we calculate a S^P (and $S^{P,SST}$) for the
Pliocene and compare this to the late Pleistocene S^P (and $S^{P,SST}$), we
also avoid complications caused by independent changes in boundary
conditions (such as topographic changes)[39] because we assess sensitiv-
ity within each relatively short time window (2.3–3.3 Myr ago versus
0–0.8 Myr ago). In addition, our approach emphasizes relative changes
in CO_2 levels and temperature over the intervals considered, rather than
absolute values. This improves accuracy because relative changes are
much better constrained than absolute temperature and $p_{CO_2}^{atm}$ values
from proxy data (see Methods for further discussion).

Preliminary regression of ΔMAT against Pliocene $p_{CO_2}^{atm}$ identified
one data point (at 2,362 kyr; white circle in Figs 1d and 2) with a par-
ticularly large residual and notable leverage on the least-squares regres-
sion (a high Cook's distance). With interglacial-like $p_{CO_2}^{atm}$ values but
glacial-like $\delta^{18}O$ values (Fig. 2), this point may reflect a chronological
error, or a short period of unusually high air–sea disequilibrium with
respect to CO_2 at ODP Site 999. To avoid the influence of this one point
on subsequent linear regressions, we have removed it from our $\delta^{11}B$–
$p_{CO_2}^{atm}$ record. The remaining $p_{CO_2}^{atm}$ data (73 points) were interpo-
lated to a constant resolution (1 kyr), smoothed with a 20-kyr moving
average to reduce short-term noise and resampled back to the original
data spacing (about one sample every 13 kyr). A Monte Carlo approach
was followed to determine uncertainties for this smoothed record given
the uncertainty in the $\delta^{11}B$-derived $p_{CO_2}^{atm}$. Radiative forcing changes
due to $p_{CO_2}^{atm}$ changes are calculated using $\Delta F_{CO2} = 5.35\ln(CO_2/C_o)$,
where $C_o = 278$ µatm (Fig. 3)[43]. We ignore mean annual forcing by
orbital variations because it is small (<0.5 W m^{-2} with a periodicity of
100–400 kyr)[31,32] and averages out over the length of our records. Linear
regressions of ΔMAT and SST$_{st}$ versus ΔF_{CO2} were performed using an
approach that yields a probabilistic estimate of slope, and hence sensi-
tivity to CO_2 forcing ($S_{CO2} = \Delta T/\Delta F_{CO2}$ or $S_{CO2,LI} = \Delta T/\Delta F_{CO2,LI}$),
which fully accounts for uncertainties in both x and y variables (see
Methods and Fig. 4). Figure 5a–d displays probability density functions
of the determinations of slope for each time interval. This analysis reveals
that, irrespective of the global temperature record used (ΔMAT or the
SST$_{st}$), the average global sensitivity of Earth's climate to forcing by
CO_2 only (S_{CO2}) is approximately twice as high for the Pleistocene as it
is for the Pliocene (Figs 4 and 5). This validates previous inferences of
a strong additional feedback factor during the Pleistocene (at $p_{CO_2}^{atm}$
levels below ~280 µatm), which probably arises from the growth and
retreat of large Northern Hemisphere ice sheets and their role in chang-
ing global albedo[31,32].

Given that, to a first order, the Earth system responds to radiative forc-
ing in a consistent fashion, largely independent of the nature of that
forcing[8], we can determine the climate forcing arising from continental
ice albedo changes via a relatively simple parameterization of sea-level
change (ΔF_{LI} = sea-level change (in units of metres) × 0.0308 W m^{-2};
following refs 31 and 32). Several reconstructions of sea-level change
partially or completely span the last 3.5 Myr (for example, refs 13, 35,
44, 45 and 46 recalculated by ref. 12), and we explore the implications
of each of these independent records. Cross-plots of combined CO_2 and
ice-albedo forcing ($\Delta F_{CO2} + \Delta F_{LI} = \Delta F_{CO2,LI}$) versus ΔMAT and SST$_{st}$
are shown in Fig. 4 for the Pliocene and Pleistocene. Figure 5e–h dis-
plays the influence of choices of temperature and sea-level record on our
determinations of $S_{CO2,LI}$ ($= \Delta T/\Delta F_{CO2,LI}$). In contrast to S_{CO2}, $S_{CO2,LI}$
is similar for both the Pliocene and Pleistocene, regardless of temper-
ature record or other parameter choices (Fig. 5). This robustly indicates
that the apparent difference between Pliocene and Pleistocene climate
sensitivity arises almost entirely from ice-albedo feedback influences.

It also implies that all of the other feedbacks that amplify climate forc-
ing by CO_2 (for example, sea-ice and cloud albedo, water vapour, vege-
tation, aerosols, other greenhouse gases) must have operated with similar
efficiencies during both the Pliocene and Pleistocene. Thus, we find no
evidence that additional (unexpected) positive feedbacks had become
active to amplify Earth system sensitivity to CO_2 forcing during the warm
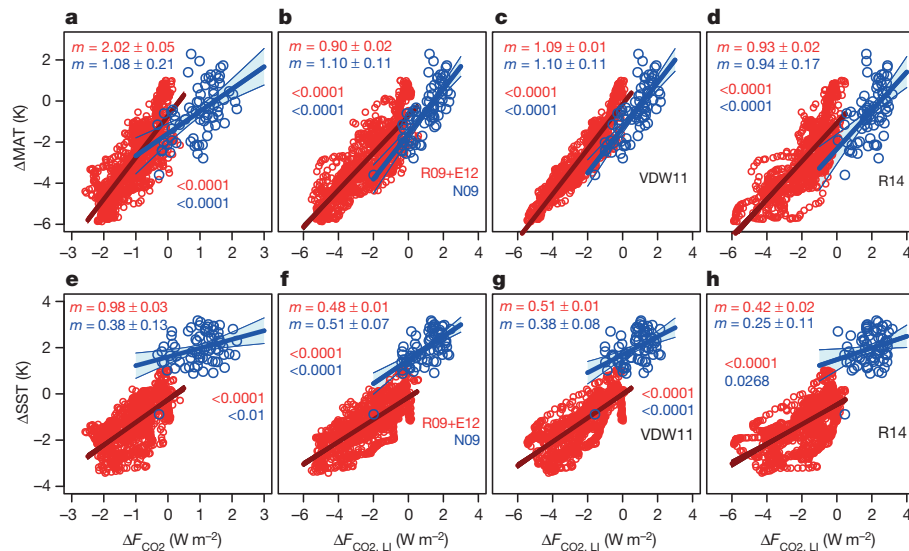Pliocene. Alternatively, if additional positive feedbacks did become active

Figure 4 | Cross plots of forcing and temperature response. a, ΔMAT versus ΔF_{CO2}, and **b–d**, ΔMAT versus $\Delta F_{CO2,LI}$, for the following sea-level records detailed in the caption for Fig. 3: **b**, R09+E12[44,45] and N09[12,46]; **c**, VDW11[35]; and **c**, R14[13]. **e**, ΔSST versus ΔF_{CO2} and **f–h**, ΔSST versus $\Delta F_{CO2,LI}$ for the same sea-level records as in **b–d**. In all panels, late Pleistocene data (0–800 kyr) are shown as red open circles and late Plio-Pleistocene (2,300–3,300 kyr) as blue open circles. Regression lines fitted by least-squares regression are also shown in the appropriate colour (shaded bands represent 95% confidence intervals). For **a–d** the temperature record is that of ref. 35 and for **e–h** it is the SST stack from this study. In all cases the slope m and standard error uncertainty are determined by least-squares regression. Also shown are the P values for the regressions.

(for example, an increase in steady-state methane concentration or changes in cloud properties), then their effect must have been negated by the loss of other amplifying feedbacks (for example, Arctic sea-ice) or the addition of more negative feedbacks. This finding is at odds with previous studies (such as refs 16,47), most probably because of differences in our approach to determine Pliocene climate sensitivity (that is, we determine a sensitivity within the Pliocene) and shortcomings in the proxy systems used by the earlier investigations, both in terms of CO_2 and temperatures (see ref. 48). For instance, Fig. 1d (and Extended Data Fig. 2) indicate that both orbital-scale variability in $p_{CO_2}^{atm}$ and the major decline at 2.7–2.9 Myr ago are absent from the previously used[16] alkenone-based $p_{CO_2}^{atm}$ records and as a result regressions of temperature and alkenone-derived forcing are poorly defined (Extended Data Fig. 2d–f).

Constraints on climate sensitivity

Using the geological record to estimate ECS directly (and thus S^a) is problematic because information on the appropriate magnitude of a number of key feedbacks (such as vegetation albedo) is typically unavailable[3]. Nonetheless, considerable effort has determined that ECS estimates based on the last glacial maximum fall within the range of ECS estimates from other approaches (1.5–4.5 K per CO_2 doubling, or 0.4–1.2 K W^{-1} m^2; ref. 1). Our analysis implies that a similar ECS applies to the Pliocene and early Pleistocene (2.3–3.3 Myr ago; Fig. 5; Supplementary Table 4). In addition, our estimate of Pliocene S_{CO2} using ΔMAT lies within a range of 0.6–1.5 K W^{-1} m^2 (at 95% confidence), meaning that, once all feedbacks have played out for future CO_2 doubling, ESS ($= S_{CO2} \times 3.7$) will very probably (95% confidence) be <5.2 K and will
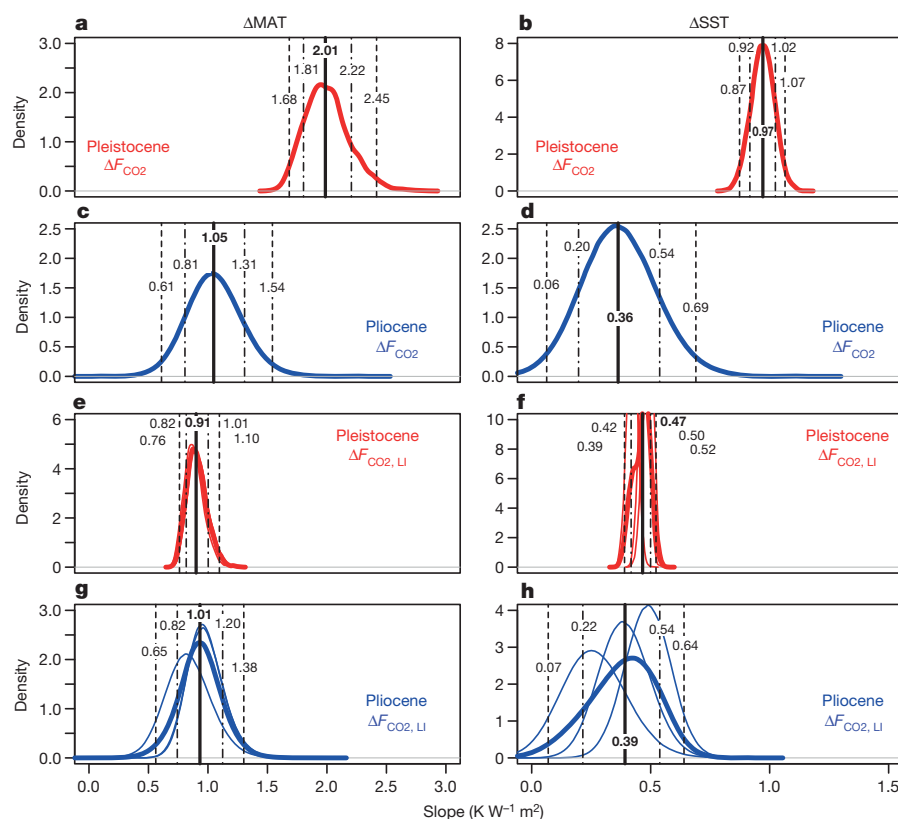

Figure 5 | Probability density functions of the slope from regressions of temperature against climate forcing. a, c, e, g, ΔMAT and **b, d, f, h**, ΔSST against ΔF_{CO2} and $\Delta F_{CO2,LI}$ for the Pleistocene (**a, b, e, f**) and Pliocene (**c, d, g, h**), taking into account the uncertainties on all variables (see text). In **e–h** individual probability density functions are shown for different choices of sea level: the combined probability density function shown in bold is the sum of these different probability density functions and therefore also incorporates uncertainty related to the choice of sea-level record. Also shown are the median (in boldface), the 68th percentile (dot-dash) and 95th percentiles (dashed).

probably (68% confidence) fall within a range of 3.0–4.4 K (Supplementary Table 4).

In May 2013, atmospheric CO_2 levels crossed the 400 parts per million threshold to values last seen during the Pliocene (Fig. 1c). Given current CO_2 emission rates, global temperatures may reach those typical of the warm periods of the Pliocene by 2050[1]. Our findings suggest that, if the Earth system behaves in a similar fashion to how it did during the Pliocene as it continues to warm in the coming years, an ECS of 1.5–4.5 K per CO_2 doubling[1] probably provides a reliable description of the Earth's equilibrium temperature response to climate forcing, at least for global temperature rise up to 3 K above the pre-industrial level. Studies of even warmer intervals in the deeper geological past (well before 3.3 Myr ago) are needed to determine whether any additional climate feedbacks should be expected as Earth warms even further into the twenty-second century if CO_2 emissions continue unabated.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1–1535 (Cambridge Univ. Press, 2013).
2. Arrhenius, S. On the influence of carbonic acid in the air upon the temperature of the ground. *Philos. Mag.* **41**, 237–276 (1896).
3. Rohling, E. J. et al. Making sense of palaeoclimate sensitivity. *Nature* **491**, 683–691 (2012).
4. Crucifix, M. Does the Last Glacial Maximum constrain climate sensitivity? *Geophys. Res. Lett.* **33**, L18701 (2006).
5. Caballero, R. & Huber, M. State-dependent climate sensitivity in past warm climates and its implication for future climate projections. *Proc. Natl Acad. Sci.* **110**, 14162–14167 (2013).
6. Beerling, D. J., Fox, A., Stevenson, D. S. & Valdes, P. J. Enhanced chemistry-climate feedbacks in past greenhouse worlds. *Proc. Natl Acad. Sci. USA* **108**, 9770–9775 (2011).
7. Meraner, K., Mauritsen, T. & Voigt, A. Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.* **40**, 5944–5948 (2013).
8. Hansen, J. et al. Efficacy of climate forcings. *J. Geophys. Res.* **110**, http://dx.doi.org/10.1029/2005JD005776 (2005).
9. Byrne, B. & Goldblatt, C. Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophys. Res. Lett.* **41**, 152–160 (2014).
10. Lunt, D. J. et al. Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geosci.* **3**, 60–64 (2010).
11. Haywood, A. M. & Valdes, P. J. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. *Earth Planet. Sci. Lett.* **218**, 363–377 (2004).
12. Miller, K. G. et al. High tide of the warm Pliocene: implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012).
13. Rohling, E. J. et al. Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508**, 477–482 (2014).
14. Lunt, D. J. et al. On the causes of mid-Pliocene warmth and polar amplification. *Earth Planet. Sci. Lett.* **321–322**, 128–138 (2012).
15. Haywood, A. M. et al. Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project. *Clim. Past* **9**, 191–209 (2013).
16. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geosci.* **3**, 27–30 (2010).
17. Beerling, D. J. & Royer, D. L. Convergent Cenozoic CO_2 history. *Nature Geosci.* **4**, 418–420 (2011).
18. Henehan, M. J. et al. Calibration of the boron isotope proxy in the planktonic foraminifera *Globigerinoides ruber* for use in palaeo-CO_2 reconstruction. *Earth Planet. Sci. Lett.* **364**, 111–122 (2013).
19. Hönisch, B. & Hemming, N. G. Surface ocean pH response to variations in pCO$_2$ through two full glacial cycles. *Earth Planet. Sci. Lett.* **236**, 305–314 (2005).
20. Foster, G. L. Seawater pH, pCO$_2$ and [CO_3^{2-}] variations in the Caribbean Sea over the last 130 kyr: a boron isotope and B/Ca study of planktic foraminifera. *Earth Planet. Sci. Lett.* **271**, 254–266 (2008).
21. Takahashi, K. et al. Climatological mean and decadal change in surface ocean pCO$_2$, and net sea-air CO_2 flux over the global oceans. *Deep Sea Res. II* **56**, 554–577 (2009).
22. Lisiecki, L. E. & Raymo, M. E. A. Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. *Paleoceanography* **20**, http://dx.doi.org/10.1029/2004PA001071 (2005).
23. Bartoli, G., Hönisch, B. & Zeebe, R. Atmospheric CO_2 decline during the Pliocene intensification of Northern Hemisphere Glaciations. *Paleoceanography* **26**, PA4213 (2012).
24. Davis, C. V., Badger, M. P. S., Bown, P. R. & Schmidt, D. N. The response of calcifying plankton to climate change in the Pliocene. *Biogeosciences* **10**, 6131–6139 (2013).
25. Seki, O. et al. Alkenone and boron based Plio-Pleistocene pCO$_2$ records. *Earth Planet. Sci. Lett.* **292**, 201–211 (2010).
26. Badger, M. P. S., Schmidt, D. N., Mackensen, A. & Pancost, R. D. High resolution alkenone palaeobarometry indicates relatively stable pCO$_2$ during the Pliocene (3.3 to 2.8 Ma). *Phil. Trans. R. Soc. A* **373**, 20130094 (2013).
27. Petit, J. R. et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
28. Lüthi, D. et al. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
29. Siegenthaler, U. et al. Stable carbon cycle-climate relationship during the Late Pleistocene. *Science* **310**, 1313–1317 (2005).
30. Hönisch, B., Hemming, G., Archer, D., Siddal, M. & McManus, J. Atmospheric carbon dioxide concentration across the Mid-Pleistocene Transition. *Science* **324**, 1551–1554 (2009).
31. Köhler, P. et al. What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Quat. Sci. Rev.* **29**, 129–145 (2010).
32. Rohling, E. J., Medina-Elizalde, M., Shepherd, J. G., Siddall, M. & Stanford, J. D. Sea surface and high-latitude temperature sensitivity to radiative forcing of climate over several glacial cycles. *J. Clim.* **25**, 1635–1656 (2012).
33. DeConto, R. M. et al. Thresholds for Cenozoic bipolar glaciation. *Nature* **455**, 652–656 (2008).
34. Shackleton, N. Oxygen isotope analyses and Pleistocene temperatures re-assessed. *Nature* **215**, 15–17 (1967).
35. van de Wal, R. S. W., de Boer, B., Lourens, L. J., Köhler, P. & Bintanja, R. Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7**, 1459–1469 (2011).
36. Balco, G. & Rovey, C. W., II. Absolute chronology for major Pleistocene advances of the Laurentide Ice Sheet. *Geology* **38**, 795–798 (2010).
37. Bailey, I. et al. An alternative suggestion for the Pliocene onset of major northern hemisphere glaciation based on geochemical provenance of North Atlantic Ocean ice-rafted debris. *Quat. Sci. Rev.* **75**, 181–194 (2013).
38. Hidy, A. J., Gosse, J. C., Froese, D. G., Bond, J. D. & Rood, D. H. A latest Pliocene age for the earliest and most extensive Cordilleran Ice Sheet in northwestern Canada. *Quat. Sci. Rev.* **61**, 77–84 (2013).
39. Lunt, D. J., Foster, G. L., Haywood, A. M. & Stone, E. J. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels. *Nature* **454**, 1102–1105 (2008).
40. Dowsett, H. J. et al. Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. *Nature Clim. Change* **2**, 365–371 (2012).
41. Shakun, J. D. et al. Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484**, 49–54 (2012).
42. Williams, R. G., Goodwin, P., Ridgwell, A. & Woodworth, P. L. How warming and steric sea level rise relate to cumulative carbon emissions. *Geophys. Res. Lett.* **39**, http://dx.doi.org/10.1029/2012GL052771 (2012).
43. Myhre, G., Highwood, E. J., Shine, K. P. & Stordal, F. New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.* **25**, 2715–2718 (1998).
44. Rohling, E. J. et al. Antarctic temperature and global sea level closely coupled over the past five glacial cycles. *Nature Geosci.* **2**, 500–504 (2009).
45. Elderfield, H. et al. Evolution of ocean tempeature and ice volume through the Mid-Pleistocene climate transition. *Science* **337**, 704–709 (2012).
46. Naish, T. R. & Wilson, G. S. Constraints on the amplitude of Mid-Pliocene (3.6–2.4 Ma) eustatic sea-level fluctuations from the New Zealand shallow-marine sediment record. *Phil. Trans. R. Soc. A* **367**, 169–187 (2009).
47. Fedorov, A. V. et al. Patterns and mechanisms of early Pliocene warmth. *Nature* **496**, 43–49 (2013).
48. O'Brien, C. L. et al. High sea surface temperatures in tropical warm pools during the Pliocene. *Nature Geosci.* **7**, 606–611 (2014).
49. Zhang, Y. G., Pagani, M., Liu, Z., Bohaty, S. M. & DeConto, R. M. A 40-million-year history of atmospheric CO_2. *Phil. Trans. R. Soc. A* **371**, 20130096 (2013).
50. Kürschner, W. M., van der Burgh, J., Visscher, H. & Dilcher, D. L. Oak leaves as biosensors of late Neogene and early Pleistocene paleoatmospheric CO_2 concentrations. *Mar. Micropaleontol.* **27**, 299–312 (1996).

Author Contributions M.A.M.-B. and T.B.C. collected the data and G.L.F. performed all relevant calculations. P.F.S. helped with sample preparation for δ^{11}B analysis and refined the age models used for ODP Sites 999 and 662. G.L.F., M.A.M.-B. and T.B.C. constructed the first draft of the manuscript and all authors contributed specialist insights that helped refine the manuscript. G.L.F., R.D.P., D.J.L. and D.N.S. conceived the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.L.F. (gavin.foster@noc.soton.ac.uk).

METHODS

Sample locations. We present new data from two deep ocean sites: ODP Site 999 (Caribbean Sea, 12° 44.64′ N and 78° 44.36′ W) and ODP Site 662 (equatorial Atlantic, 1° 23.41′ S, 11° 44.35′ W). Both sites have well-constrained age models for the Pliocene and are part of the Lisiecki and Raymo benthic foraminifera δ^{18}O stack[22] (hereafter LR04). Sedimentation rates are comparable between the sites ($\sim$3 cm kyr^{-1} at ODP 999 and $\sim$4 cm kyr^{-1} at ODP 662). At ODP Site 999, seventy-four samples were analysed at an average temporal resolution of around one sample every 13 kyr, targeting several glacial and interglacial maxima. ODP Site 662 was analysed at much lower resolution (8 samples in 1,000 kyr = 1 sample every 125 kyr on average), and the chosen samples were limited to peak interglacial conditions to avoid potential upwelling influences during glacial periods[51]. The extent of the modern air–sea CO_2 disequilibrium at each location is displayed in Extended Data Fig. 1a.

Analytical methodology. Between 90 and 200 individuals of *G. ruber* ($\sim$10 µg per shell) were picked from the 300–355-µm size fraction from ODP Sites 999 and 662. Foraminiferal samples were crushed between cleaned glass microscope slides and subsequently cleaned according to established oxidative cleaning methods[52–54]. After cleaning, samples were dissolved in $\sim$0.15 M Teflon-distilled HNO$_3$, centrifuged and transferred to 5 ml Teflon vials for storage. An aliquot ($\sim$20 µl; $\sim$7% of the total sample) was taken for trace element analysis. Boron was separated from the dissolved samples using Amberlite IRA-743 boron-specific anion exchange resin following established procedures[20]. Boron isotope ratios were measured on a Thermo Scientific Neptune multicollector inductively coupled plasma mass spectrometer (MC-ICPMS) at the University of Southampton according to methods described elsewhere[18,20,54].

External reproducibility of δ^{11}B analyses is calculated following the approach of ref. 54, and is described by the relationship:

$$2\sigma = 1.87\exp^{-20.6[^{11}B]} + 0.22\exp^{-0.43[^{11}B]} \quad (1)$$

where [^{11}B] is the intensity of the ^{11}B signal in volts (see ref. 18 for further details).

Trace elements were measured on a Thermo Scientific Element 2 single collector ICPMS at the University of Southampton, following established methods[20]. Over the period 2012–2013, analytical reproducibility for Mg/Ca was $\pm$2.7% (2σ). Raw Mg/Ca ratios were corrected for changes in the Mg/Ca ratio of seawater (Mg/Ca)$_{sw}$, using the approach of ref. 55 and the power-law modification of ref. 56 and the modelled (Mg/Ca)$_{sw}$ of ref. 57. Specifically, we use an H value[56] of 0.41, originally derived for *G. sacculifer*[58], because no species-specific H value is currently available for *G. ruber* (for extended discussion, see ref. 48). The following equation[56,59] was therefore used to derive calcification temperatures (in units of °C) from our Mg/Ca ratios, which also includes a depth correction to account for the influence of dissolution on shell Mg/Ca ratios.

$$T = \frac{\ln\left[\left(\frac{Mg}{Ca}\right)_{test} / E\left(\left(\frac{Mg}{Ca}\right)^t_{sw}\right)^H\right]}{0.09} + (0.61Z) \quad (2)$$

where $\left(\frac{Mg}{Ca}\right)^t_{sw}$ is the Mg/Ca ratio of seawater at the time t of interest, $\left(\frac{Mg}{Ca}\right)_{test}$ is the Mg/Ca of the foraminiferal test, Z is the core depth in kilometres and E is defined by the following equation[56]:

$$E = \frac{0.38}{\left(\left(\frac{Mg}{Ca}\right)^t_{sw}\right)^H} \quad$$

Trace element data were also used to check the efficiency of the foraminiferal cleaning procedure[20,54]. All samples had Al/Ca ratios of <100 µmol mol^{-1}, and typically <60 µmol mol^{-1}.

Determination of pH from δ^{11}B of *G. ruber*. Boron in seawater exists mainly as two different species, boric acid (B(OH)$_3$) and the borate ion (B(OH)$_4^-$), and their relative abundance is pH dependent. There are two isotopes of boron, ^{11}B ($\sim$80%) and ^{10}B ($\sim$20%), with a ratio normally expressed in delta notation (in per mil, ‰) as:

$$\delta^{11}B = \left[\left(\frac{^{11}B/^{10}B_{sample}}{^{11}B/^{10}B_{NIST951}}\right) - 1\right] \times 1,000 \quad (3)$$

where ^{11}B/^{10}B$_{NIST951}$ is the isotopic ratio of the NIST SRM 951 boric acid standard (^{11}B/^{10}B = 4.04367; ref. 60).

There is a pronounced isotopic fractionation between the two dissolved boron species, with boric acid being enriched in ^{11}B by 27.2‰ (ref. 61). As the concentration of each species is pH dependent, their isotopic composition also has to change with pH in order to maintain a constant seawater δ^{11}B. Calibration studies[54,62,63] have shown that the borate species is predominantly incorporated into foraminiferal

CaCO$_3$, and therefore ocean pH can be calculated from the δ^{11}B of borate (δ^{11}B$_{borate}$) as follows:

$$pH = pK_B^* - \log\left(-\frac{\delta^{11}B_{sw} - \delta^{11}B_{borate}}{\delta^{11}B_{sw} - \left(^{11,10}K_B \times \delta^{11}B_{borate}\right) - 1,000\left(^{11,10}K_B - 1\right)}\right) \quad (4)$$

where pK_B^* is the dissociation constant for boric acid at *in situ* temperature, salinity and pressure[64], δ^{11}B$_{sw}$ is the isotopic composition of seawater (39.61‰; ref. 65), δ^{11}B$_{borate}$ is the isotopic composition of borate ion, and $^{11,10}K_B$ is the isotopic fractionation between the two aqueous species of boron in seawater (1.0272 $\pm$ 0.0006) (ref. 61).

In our calculations, the temperature for ODP Site 999 is derived from Mg/Ca ratios measured on aliquots (separated after dissolution) of the same samples as those used for δ^{11}B analysis and for ODP Site 662 from published records of temperature using the U$^{K'}_{37}$ proxy[66]. Despite the uncertainty in Mg/Ca-derived SSTs we have not used published U$^{K'}_{37}$ temperature records for ODP Site 999 because they are of lower temporal resolution and close to saturation ($T = 28$–29 °C)[25]. Salinity has little influence on the calculations of pH ($\pm$1 psu = $\pm$0.006 pH units), and therefore is assumed to be constant at 35 psu (similar to the present-day mean annual average at both locations). The uncertainty associated with this assumption is propagated into $p_{CO_2}^{atm}$ calculations.

Boron has a long residence time in seawater (10–20 Myr; ref. 67), and to account for likely (small) changes in the boron isotopic composition of seawater (δ^{11}B$_{sw}$) over the last 3 Myr, we use a simple linear extrapolation between modern δ^{11}B$_{sw}$ (39.61‰; ref. 65) and the δ^{11}B$_{sw}$ determined by ref. 68 for the middle Miocene (12.72 Myr ago; δ^{11}B$_{sw}$ = 37.8‰). This simple estimation yields δ^{11}B$_{sw}$ = 39.2‰ at 3 Myr ago, which is consistent with available independent constraints, for example those based on assumptions of bottom-water pH and measured benthic foraminiferal δ^{11}B (ref. 69).

Finally, to calculate pH from the δ^{11}B of *G. ruber*, it is necessary to account for species-specific differences between δ^{11}B$_{borate}$ in ambient seawater and δ^{11}B in foraminiferal calcite (δ^{11}B$_{calcite}$; that is, 'vital effects'). Here we used the species- and size-specific calibration equation of ref. 18 for *G. ruber* in the size range 300–355 µm (see equation (5)). This equation has been applied in previous studies[18] to produce a δ^{11}B-based $p_{CO_2}^{atm}$ record for the last 30 kyr that is in very good agreement with ice-core $p_{CO_2}^{atm}$ records (Extended Data Fig. 2).

$$\delta^{11}B_{borate} = (\delta^{11}B_{calcite} - 8.87 \pm 1.52)/0.60 \pm 0.08 \quad \text{(uncertainty at } 2\sigma) \quad (5)$$

It is important to note that not only is there generally good preservation of the sites we use[23,24], but also the δ^{11}B of *G. ruber* does not appear to be greatly affected by partial dissolution[25].

Determination of $p_{CO_2}^{atm}$ from δ^{11}B-derived pH. Another variable of the ocean carbonate system is required besides pH to calculate the partial pressure of CO_2 in seawater, $p_{CO_2}^{sw}$ (ref. 70). Here, total alkalinity (TA) is assumed to be constant at values similar to modern values at ODP Site 999 (2,330 µmol kg^{-1}; ref. 20). It is important to note that $p_{CO_2}^{sw}$ estimates are mostly determined by the reconstructed pH and that TA has little influence. This is because pH reflects the ratio of TA to DIC (total dissolved inorganic carbon), so when pH is known the ratio of TA:DIC is set, so the effect on $p_{CO_2}^{sw}$ of a large increase/decrease in TA is partially countered by an opposite change in DIC. Indeed, at a given pH, a 10% change in TA results in a $p_{CO_2}^{sw}$ change of only 10%. For example, modifying TA by $\pm$100 µmol kg^{-1} (a range equivalent to modelled variations in TA for the last 2 Myr; ref. 30) modifies reconstructed $p_{CO_2}^{sw}$ (when pH is known) by less than $\pm$12 µatm.

$p_{CO_2}^{sw}$ was calculated using the equations of ref. 70, the "seacarb" package of R (statistical software, see ref. 71) and a Monte Carlo approach ($n = 10,000$) to fully propagate the uncertainty in the input parameters (at 95% confidence or full range, where appropriate): δ^{11}B ($\pm$analytical uncertainty, calculated using equation (1), and the calibration uncertainty in equation (5)), the Mg/Ca-derived temperature ($\pm$3 °C), the salinity ($\pm$3 psu), TA ($\pm$175 µmol kg^{-1}), and δ^{11}B$_{sw}$ ($\pm$0.4‰). $p_{CO_2}^{atm}$ was then calculated from $p_{CO_2}^{sw}$ using Henry's Law and subtracting the modern disequilibria with respect to CO_2 at the two sites (Extended Data Fig. 1; Supplementary Tables 1 and 2). Note that for the quoted uncertainty range for temperature, salinity, and δ^{11}B$_{sw}$ a normal distribution is assumed. However, for TA we have assumed a 'flat' probability (that is, an equal probability of TA being any value in the range 2,155–2,505 µmol kg^{-1}). We therefore do not ascribe weight to the assumption that TA remains constant, but rather fully explore the likely range given the available, model-based, constraints[72,73]. It should also be noted that salinity and temperature have little effect on our estimated $p_{CO_2}^{sw}$ ($+$1 psu = $+$0.2 µatm; $+$1 °C = $+$8 µatm).

Comparison with published records of Pliocene $p_{CO_2}^{atm}$. Figure 1 and Extended Data Fig. 2b, c show a comparison of our high-resolution δ^{11}B-derived $p_{CO_2}^{atm}$ record with published records. As noted in the main text, although the various approaches agree, in detail our record exhibits more structure. As a consequence, cross plots of

the previously published CO_2 data against ΔMAT (or SST_{st}) are largely incoherent (Extended Data Fig. 2d–f). In the case of the stomatal estimates[50] and the existing δ^{11}B-based records[23,25], this is mainly a consequence of their low temporal resolution, although analytical issues[74] and species choice (we use *G. ruber*, which spends its entire life cycle in the mixed layer, whereas ref. 23 uses *G. sacculifer*, which migrates during its life cycle and whose δ^{11}B, unlike that of *G. ruber*, is modified by partial dissolution[25]) may also have a role in the discrepancy with earlier δ^{11}B records (see ref. 25 for further discussion). The lack of variability through the Pliocene for the alkenone-based records may be related to changes in the size of the alkenone producers[26], fluctuations in nutrient content/water depth of maximum alkenone production, and/or variations in the degree of passive versus active uptake of CO_2 by the alkenone-producing coccolithophorids[49,75].

Continuous records of Plio-Pleistocene global temperature change. Robust records of global temperature change are needed to determine how Earth's climate has responded to changes in CO_2. Here we estimate this variable using two independent approaches: (1) we generate a stack of available sea surface temperature records (SST_{st}); and (2) following ref. 35 we use a reconstruction of global mean annual surface air temperature change based on a scaling of the Northern Hemisphere temperature required by a simple coupled ice-sheet–climate model to predict the benthic δ^{18}O stack of ref. 76 (tuned here to the LR04 age model; ΔMAT).

For the SST stack we imposed a number of criteria for site selection. These are: (1) the record must be continuous from the late Pliocene to the late Pleistocene (or nearly so); (2) the temporal resolution must be relatively high (ideally better than one sample per 10 kyr; for ODP Site 1237 we have, however, accepted a lower resolution to increase spatial coverage) to allow us to fully resolve the dominant orbital-scale variability; (3) the SST record must be based on $U^{K'}_{37}$, given that Mg/Ca suffers an unacceptable level of uncertainty on these timescales owing to the secular evolution of the Mg/Ca ratio of seawater (for example, ref. 48); and (4) the temperatures recorded by the $U^{K'}_{37}$ proxy must be less than 29 °C, above which the proxy becomes saturated and therefore unresponsive[92]. Ten published records meet these criteria (ODP Sites 982, 607, 1012, 1082, 1239, 846, 662, 722, 1237 and 1090; refs 66, 77–84) and the locations of these sites are shown in Extended Data Fig. 2b. The average temporal resolution of these records is one sample about every 4 kyr (ranging from ~2 kyr to ~13 kyr) and the published age model of each site is either part of the LR04 stack or was tuned to it (see the original publications for details).

To stack the records, each was first converted to a relative SST record referenced to either the average of the Holocene (0–10 kyr), or mean annual modern SST if the Holocene is missing, and then linearly interpolated to a 5-kyr spacing. These relative records are then averaged to produce a single stacked record of relative SST change (SST_{st}; Supplementary Table 5). The number of sites contributing to the SST stack varies but for most of the record is $\geq$8 (Extended Data Fig. 6a, b). Uncertainty on the SST stack is estimated by a Monte Carlo procedure where 1,000 realizations are made of each individual SST record with noise added reflecting the magnitude of analytical uncertainty in the $U^{K'}_{37}$ SST reconstruction (±1 °C at 2σ; ref. 92). Since we are using the same proxy for each location it is not necessary to consider the calibration uncertainty, as this should be the same for each record. Each SST realization is then averaged to produce 1,000 realizations of the SST stack. The mean of these 1,000 realizations is then calculated and the 95% confidence interval is given by the 2.5% and 97.5% percentile (red band on Fig. 3). Jacknifing of the SST stack (that is, the sequential removal of one record at a time) indicates that no particular record has undue influence and the SST stack remains close to the bounds relating to analytical uncertainty (the grey lines on Extended Data Fig. 6c, d).

Our aim with the SST stack was not specifically to reconstruct global SST change but rather to examine the change in SST at these locations for a given forcing in the Pliocene and Pleistocene. We therefore do not require the SST stack to reflect global SST change. However, in order to assess how well the SST stack does reflect global SST we: (1) Examined the mean of historic SST change (1870 AD to 2013 AD; from the HadISST database; ref. 85) at each location where we have an alkenone palaeo-SST record. This comparison is shown in Extended Data Fig. 7 (blue circles). Despite exhibiting more variability than the mean annual global average (red in Extended Data Fig. 7), these ten sites clearly capture the global long-term trend in global mean SST[86,87] over the last 140 years or so (Extended Data Fig. 7). (2) Compared the SST stack to a multi-proxy and more comprehensive and independent compilation of ref. 32 that covers the last 100 kyr with >30 sites and the last 278 kyr with >10 sites. When data for the last 278 kyr are stacked together in a similar way to the SST stack, the stack of ref. 32 (blue on Fig. 3c) compares well with SST_{st}, giving us confidence that it closely reflects global SST change. (3) Compared the SST stack to discrete global reconstructions of SST. For the last glacial (20–25 kyr), the SST stack gives a ΔSST of -2.2 ± 0.4 K, which is close to the ΔSST of -3.2 K from a recent comprehensive compilation for the Last Glacial Maximum[41] and is within uncertainty of earlier reconstructions (for example, ref. 91 where ΔSST of -1.9 ± 1.8 K). For the Mid-Pliocene Warm Period (3–3.3 Myr ago), our SST stack gives an average of $+2.3$ K. A simple mean calculated from the larger multi-proxy

PRISM SST compilation of ref. 40 is very similar at $+2.6$ K. The SST stack is slightly warmer than an area-weighted mean of the PRISM SST set ($+2$ K; ref. 40).

Taken together, these comparisons clearly indicate that, although SST_{st} is made of a limited number of sites, it does appear to closely reflect change in global SST. This conclusion is also supported by the general agreement between the trends (but not absolute values) exhibited by ΔMAT and the SST stack through the Pliocene and Pleistocene (Fig. 3), with subtle differences between these two climate records (for example, at 2.8 Myr ago) potentially a result of a decoupling between deep- and surface-water temperature evolution, small spatial biases in our SST stack, and/or minor age-model inaccuracies (the conversion of depth below seafloor in a marine sediment core to age).

Regression-based determinations of climate sensitivity. To examine the climatic response (expressed as either ΔMAT or ΔSST) to forcing by CO_2 and land-ice albedo changes in both time periods, we used a linear regression approach. Because each variable used (CO_2 and sea level, ΔMAT or ΔSST) has an associated uncertainty, however, it is necessary to fully explore the influence of these uncertainties on our estimates of slope determined using least-squares linear regression. Owing to the difficulty of performing the least-squares linear regression with uncertainty in x and y variables that are not necessarily normally distributed, we have used a two-stage approach to fully propagate all the uncertainties involved. First, we generated 1,000 realizations of each temporal record of each variable (for example, ΔF_{CO_2}, $\Delta F_{CO2,LI}$, ΔMAT or ΔSST) based on a random sampling of each record within its uncertainty envelope. This uncertainty envelope was either a simple normal distribution (for example, ±6 parts per million for ice-core CO_2) or based on other Monte Carlo output (for example, random sampling of the 10,000 simulations of the Pliocene δ^{11}B–$p_{CO_2}^{atm}$ record or the 1,000 realizations of the SST stack; see above). Then the first realization of the ΔF_{CO_2} (or $\Delta F_{CO2,LI}$) record was regressed against the first realization of the ΔMAT (or ΔSST) with the uncertainty in the slope and intercept of that regression determined using a bootstrapping approach ($n = 1,000$; ref. 88). The second realization of the forcing term and the climate response was then regressed and the 1,000 estimates of slope and intercept by bootstrapping were combined with 1,000 of the first regression. This continued for all 1,000 realizations and a probability density function for the slope and intercept, accounting for x and y uncertainty, was then constructed from the combined bootstrap estimates for each realization ($n = 1,000,000$). The results of this approach are shown in Fig. 5.

As noted above, $p_{CO_2}^{atm}$ (and hence ΔF_{CO_2}) calculated from boron isotopes is a function of not only the measured δ^{11}B but also the total alkalinity (TA; or other second carbonate system variable) and, beyond the last million years or so, the boron isotopic composition of seawater ($\delta^{11}B_{sw}$). This is illustrated in Extended Data Fig. 8. Here $p_{CO_2}^{atm}$ is calculated from an artificial δ^{11}B and temperature record (Extended Data Fig. 8a), a TA of either 2,000 µmol kg^{-1}, 2,300 µmol kg^{-1} or 2,600 µmol kg^{-1}, a $\delta^{11}B_{sw}$ of 38.8‰, 39.6‰ (that is, modern) or 40.4‰ (Extended Data Fig. 8) and the assumption that $p_{CO_2}^{atm} = p_{CO_2}^{sw}$. These parameter choices result in a large difference in absolute CO_2 but, although they are extreme and perhaps unlikely for the Pliocene, the slope of a linear regression of global temperature change and ΔF_{CO_2} is very similar for each set of parameters (Extended Data Fig. 8c, d). So much so, that even with only a poor knowledge of $\delta^{11}B_{sw}$ (for example, ±0.8‰) and TA (for example, ±300 µmol kg^{-1}) the accuracy of the relationship between reconstructed ΔF_{CO_2} and temperature is not unduly affected.

The residence time of boron in seawater (10–20 Myr) ensures that changes in $\delta^{11}B_{sw}$ across the time interval examined here (1 Myr) are unlikely to be large (<0.1‰; ref. 67) and so uncertainty in the absolute value of $\delta^{11}B_{sw}$ and any changes across the study interval can be ignored for our determinations of S^p. In all the previous calculations we assume that TA is randomly distributed between 2,155 µmol kg^{-1} and 2,505 µmol kg^{-1}, therefore accounting for all possible trends in TA across the time interval studied within this range. However, to better examine the influence of a large secular shift in TA on our estimates of S^p we have imposed a 200 µmol kg^{-1} decrease (TA$_d$) or increase (TA$_i$) across our Pliocene study interval. The slope for the regressions using one parameter set (VDW11 and sea level values from ref. 46 recalculated by ref. 12) but with such a varying TA are shown in Extended Data Fig. 8e, f. Even this relatively large secular change does not have a major influence on the estimated slope, clearly illustrating that our assumptions regarding TA, both its absolute value and its secular evolution, have little influence on our calculated ΔF_{CO_2} and hence our conclusions.

Pliocene $p_{CO_2}^{atm}$ variability. The apparent cyclicity in our Pliocene CO_2 record can be investigated using spectral analysis. Extended Data Fig. 4c shows that the evolutive power spectra for the Pliocene $p_{CO_2}^{atm}$ and a ~100-kyr cycle is clearly dominant. Our sampling resolution is one sample per ~13 kyr, which is not sufficient to resolve cycles of a precessional length (for example, 19 kyr and 23 kyr) but may be adequate to resolve obliquity (~41-kyr length), yet these cycles are apparently absent in the generated spectra (Extended Data Fig. 4c). To ensure our resolution is not biasing this result we have sampled the LR04 benthic δ^{18}O stack at our exact sampling resolution and examined the evolutive power spectra of this

sampled record (Extended Data Fig. 4d). This analysis reveals the presence of 100-kyr and 41-kyr cycles in the $\delta^{18}O$ data, despite our relatively low resolution, supporting the observation that the dominant cycle in Pliocene $p_{CO_2}^{atm}$ is ~100 kyr.

The magnitude of Pliocene $p_{CO_2}^{atm}$ variability, shown in Extended Data Fig. 4a, is similar to that exhibited by published late and mid-Pleistocene $\delta^{11}B$–$p_{CO_2}^{atm}$ records (green and red lines on Extended Data Fig. 4a) and by the Late Pleistocene ice-core data when noise that is approximately equivalent to our $\delta^{11}B$–$p_{CO_2}^{atm}$ uncertainty is added (± 35 µatm; black dashed line on Extended Data Fig. 4a). In contrast, the $\delta^{18}O$ variability for these time intervals increases markedly from the Pliocene to late Pleistocene as the magnitude of glacial–interglacial cycles increases (Fig. 1e, Extended Data Fig. 4b).

51. Lawrence, K. T. *et al.* Time-transgressive productivity changes in the North Atlantic upon Northern Hemisphere glaciation. *Paleoceanography* **28,** 740–751 (2013).
52. Barker, S., Greaves, M. & Elderfield, H. A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. *Geochem. Geophys. Geosyst.* **4,** 8407 (2003).
53. Yu, J., Elderfield, H., Greaves, M. & Day, J. Preferential dissolution of benthic foraminiferal calcite during laboratory reductive cleaning. *Geochem. Geophys. Geosyst.* **8,** Q06016 (2007).
54. Rae, J. W. B., Foster, G. L., Schmidt, D. N. & Elliott, T. Boron isotopes and B/Ca in benthic foraminifera: proxies for the deep ocean carbonate system. *Earth Planet. Sci. Lett.* **302,** 403–413 (2011).
55. Medina-Elizalde, M., Lea, D. W. & Fantle, M. S. Implications of seawater Mg/Ca variability for Plio-Pleistocene tropical climate reconstruction. *Earth Planet. Sci. Lett.* **269,** 585–595 (2008).
56. Evans, D. & Muller, W. Deep time foraminifera Mg/Ca paleothermometry: nonlinear correction for secular change in seawater Mg/Ca. *Paleoceanography* **27,** PA4205 (2012).
57. Fantle, M. S. & DePaolo, D. J. Sr isotopes and pore fluid chemistry in carbonate sediment of the Ontong Java Plateau: calcite recrystallisation rates and evidence for a rapid rise in seawater Mg over the last 10 million years. *Geochim. Cosmochim. Acta* **70,** 3883–3904 (2006).
58. Delaney, M. L., Bé, W. H. A. & Boyle, E. A. Li, Sr, Mg, and Na in foraminiferal calcite shells from laboratory culture, sediment traps, and sediment cores. *Geochim. Cosmochim. Acta* **49,** 1327–1341 (1985).
59. Dekens, P. S., Lea, D. W., Pak, D. K. & Spero, H. J. Core top calibration of Mg/Ca in tropical foraminifera: refining paleotemperature estimation. *Geochem. Geophys. Geosyst.* **3,** http://dx.doi.org/10.1029/2001GC000200 (2002).
60. Catanzaro, E. J. *et al.* Boric Assay; Isotopic, and Assay Standard Reference Materials (US National Bureau of Standards Special Publication 260-17, 1970).
61. Klochko, K., Kaufman, A. J., Yoa, W., Byrne, R. H. & Tossell, J. A. Experimental measurement of boron isotope fractionation in seawater. *Earth Planet. Sci. Lett.* **248,** 276–285 (2006).
62. Hemming, N. G. & Hanson, G. N. Boron isotopic composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **56,** 537–543 (1992).
63. Hemming, N. G., Reeder, R. J. & Hanson, G. N. Mineral-fluid partitioning and isotopic fractionation of boron in synthetic calcium carbonate. *Geochim. Cosmochim. Acta* **59,** 371–379 (1995).
64. Dickson, A. G. Thermodynamics of the dissociation of boric-acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Res.* **37,** 755–766 (1990).
65. Foster, G. L., Pogge von Strandmann, P. A. E. & Rae, J. W. B. Boron and magnesium isotopic composition of seawater. *Geochem. Geophys. Geosyst.* **11,** Q08015 (2010).
66. Herbert, T. D., Cleaveland Peterson, L., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328,** 1530–1534 (2010).
67. Lemarchand, D., Gaillardet, J., Lewin, E. & Allegre, C. J. Boron isotope systematics in large rivers: implications for the marine boron budget and paleo-pH reconstruction over the Cenozoic. *Chem. Geol.* **190,** 123–140 (2002).
68. Foster, G. L., Lear, C. H. & Rae, J. W. B. The evolution of pCO₂, ice volume and climate during the middle Miocene. *Earth Planet. Sci. Lett.* **341-344,** 243–254 (2012).
69. Raitzsch, H. B. Cenozoic boron isotope variations in benthic foraminifers. *Geology* **41,** 591–594 (2013).
70. Zeebe, R. & Wolf-Gladrow, D. A. *CO₂ in Seawater: Equilibrium, Kinetics, Isotopes* (Elsevier Oceanography Series 65, 2001).
71. R Core Team. *R: a Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, 2013).
72. Tyrrell, T. & Zeebe, R. E. History of carbonate ion concentration over the last 100 million years. *Geochim. Cosmochim. Acta* **68,** 3521–3530 (2004).
73. Clark, P. U. *et al.* The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in pCO₂. *Quat. Sci. Rev.* **25,** 3150–3184 (2006).
74. Foster, G. L. *et al.* Interlaboratory comparison of boron isotope analyses of boric acid, seawater and marine CaCO₃ by MC-ICPMS and NTIMS. *Chem. Geol.* **358,** 1–14 (2013).
75. Bolton, C. T. & Stoll, H. M. Late Miocene threshold response of marine algae to carbon dioxide limitation. *Nature* **500,** 558–562 (2013).
76. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686–693 (2001).
77. Lawrence, K. T., Herbert, T. D., Brown, C. M., Raymo, M. E. & Haywood, A. M. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography* **24,** PA2218 (2009).
78. Lawrence, K. T., Sosdian, S., White, J. M. & Rosenthal, Y. North Atlantic climate evolution through the Plio-Pleistocene climate transitions. *Earth Planet. Sci. Lett.* **300,** 329–342 (2010).
79. Brierley, C. M. *et al.* Greatly expanded tropical warm pool and weakened Hadley Circulation in the Early Pliocene. *Science* **323,** 1714–1718 (2009).
80. Etourneau, J., Martinez, P., Blanz, T. & Schneider, R. Pliocene-Pleistocene variability of upwelling activity, productivity, and nutrient cycling in the Benguela region. *Geology* **37,** 871–874 (2009).
81. Etourneau, J., Schneider, R., Blanz, T. & Martinez, P. Intensification of the Walker and Hadley atmospheric circulations during the Pliocene-Pleistocene climate transition. *Earth Planet. Sci. Lett.* **297,** 103–110 (2010).
82. Lawrence, K. T., Liu, Z. & Herbert, T. D. Evolution of the eastern tropical Pacific through Plio-Pleistocene glaciation. *Science* **312,** 79–83 (2006).
83. Martinez-Garcia, A., Rosell-Melé, A., McClymont, E. L., Gersonde, R. & Haug, G. H. Subpolar link to the emergence of the modern equatorial Pacific cold tongue. *Science* **328,** 1550–1553 (2010).
84. Dekens, P. S., Ravelo, A. C. & McCarthy, M. D. Warm upwelling regions in the Pliocene warm period. *Paleoceanography* **22,** PA3211 22 (2007).
85. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. D* **108,** 4407 (2003).
86. Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations in situ since 1850: 1. Measurement and sampling uncertainty. *J. Geophys. Res.* **116,** D14103 (2011a).
87. Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.* **116,** D14104 (2011b).
88. Efron, B. Bootstrap methods: another look at the jacknife. *Ann. Stat.* **7,** 1–26 (1979).
89. Locarnini, R. A. *et al.* World Ocean Atlas 2013, Volume 1: Temperature (NOAA Atlas NESDIS 73, 2013).
90. Schlitzer, R. *Ocean Data View* http://odv.awi.de (2012).
91. Waelbroeck, C. *et al.* Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geosci.* **2,** 127–132 (2009).
92. Müller, P. J., Kirst, G., Ruhland, G., Von Storch, I. & Rosell-Melé, A. Calibration of the alkenone paleotemperature index UK'37 based on core-tops from the eastern South Atlantic and the global ocean (60 N-60 S). *Geochim. Cosmochim. Acta* **62,** 1757–1772 (1998).
93. Siddall, M. *et al.* Sea-level fluctuations during the last glacial cycle. *Nature* **423,** 853–858 (2003).
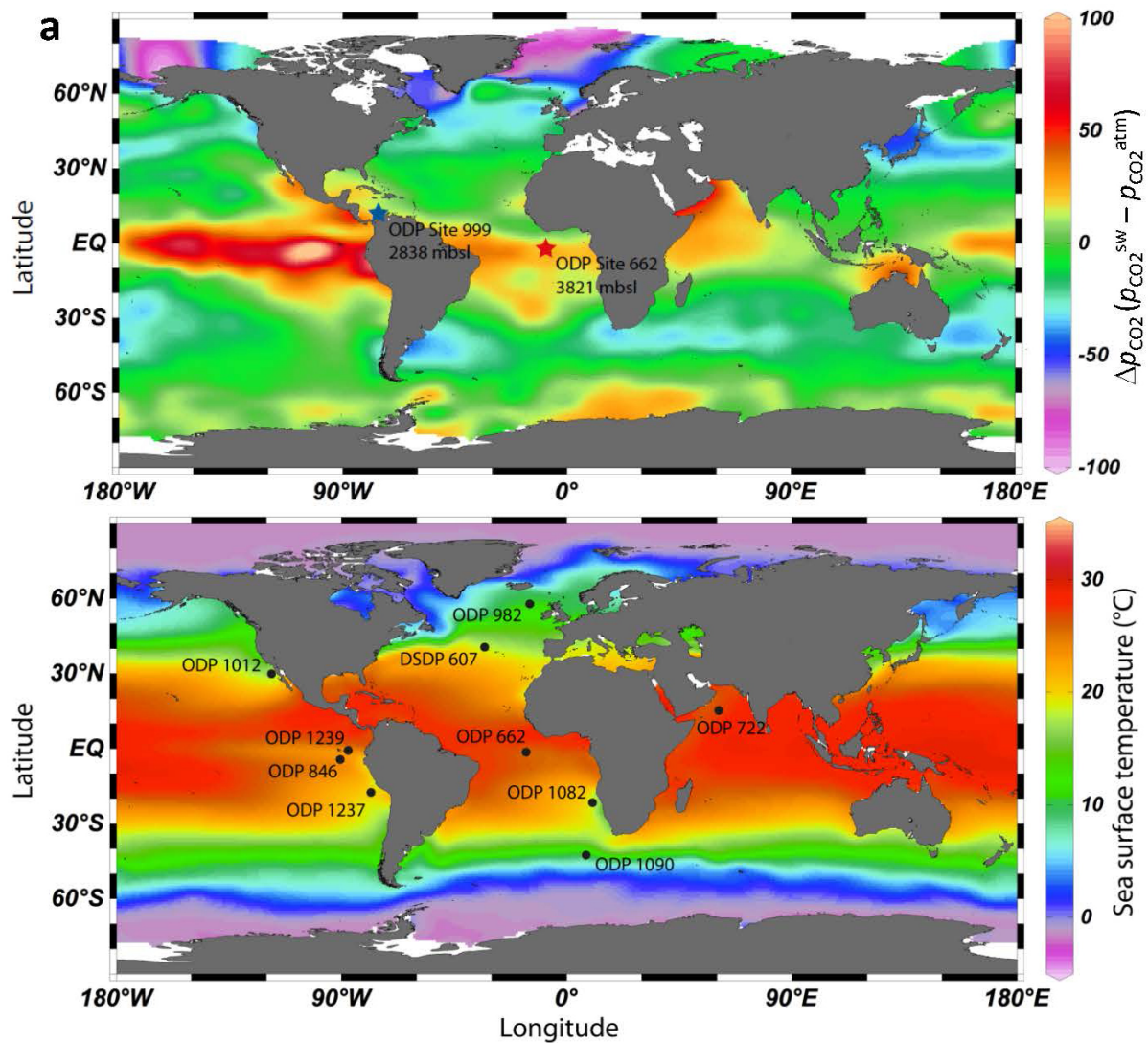
Extended Data Figure 1 | Maps of modern mean annual Δp_{CO_2} and SST labelled with site locations. **a**, Map of sites used for $p_{CO_2}^{atm}$ reconstructions with the mean annual modern Δp_{CO_2} from the reconstruction of ref. 21. **b**, Map of the sites (and labelled with their depths) used to generate the SST stack with mean annual modern SST from the World Ocean Atlas 2013 (ref. 89). mbsl, metres below sea level, where DSDP is the Deep Sea Drilling Project. Figures constructed and data visualized in Ocean Data View[90].
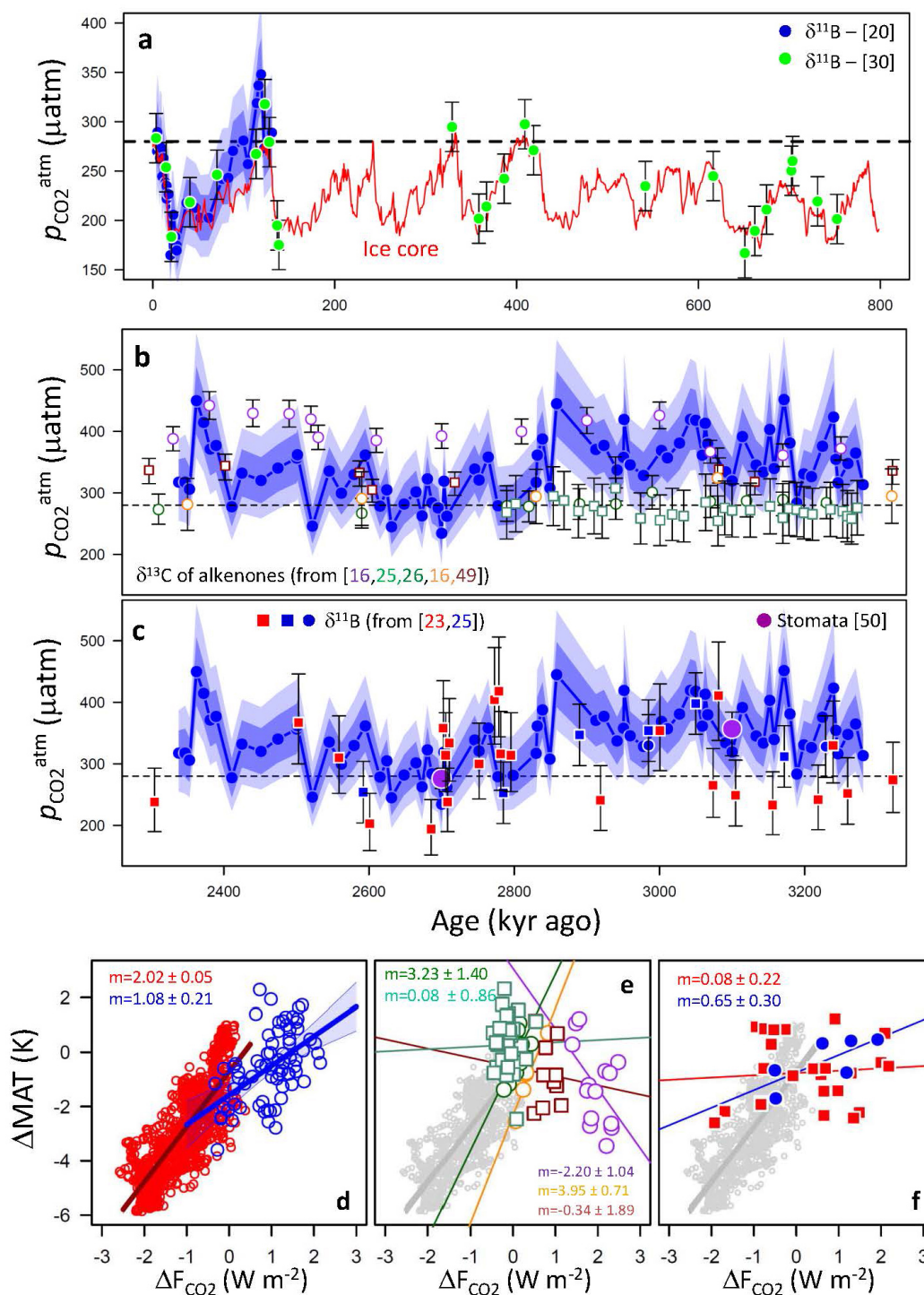
Extended Data Figure 2 | Comparisons of boron-isotope-based $p_{CO_2}^{atm}$ estimates with other methodologies and archives. **a**, Estimates of $p_{CO_2}^{atm}$ from published $\delta^{11}B$ records compared to ice-core CO_2 (red line; refs 27–29). The dotted line is for $p_{CO_2} = 278\,\mu atm$. In **a** the data of ref. 20 (blue circles) have been recalculated in the same manner as described here for the Pliocene, including using the *G. ruber* $\delta^{11}B$–pH calibration of ref. 18. The error band encompasses 68% (dark blue) and 95% (light blue) of 10,000 Monte Carlo simulations of $p_{CO_2}^{atm}$ (see main text). Also shown are the *G. sacculifer*-based $\delta^{11}B$–$p_{CO_2}^{atm}$ record of ref. 30 (green circles). In this case error bars ($\pm 25\,\mu atm$) are as determined in that study. Despite similar analytical uncertainty, the smaller error bars for the ref. 30 data result from these authors not propagating the $\delta^{11}B$–pH calibration uncertainty and considering a smaller range in temperature, salinity and alkalinity uncertainty than in this study ($\pm 0.76\,°C$, $\pm 1\,psu$, $\pm 27\,\mu mol\,kg^{-1}$ versus $\pm 3\,°C$, $\pm 3$ practical salinity units (psu), $\pm 175\,\mu mol\,kg^{-1}$ with a flat probability in this study). **b**, $\delta^{11}B$-based $p_{CO_2}^{atm}$

record generated here (blue closed circles and 95% and 68% uncertainty bands) with $p_{CO_2}^{atm}$ from the $\delta^{13}C$ of alkenones from published studies. See Fig. 1 legend for details. **c**, $\delta^{11}B$-based $p_{CO_2}^{atm}$ record generated here (blue closed circles and 95% and 68% uncertainty bands) with $p_{CO_2}^{atm}$ from previous $\delta^{11}B$-based studies and from plant stomata. See Fig. 1 legend for details. **d–f**, Comparison of cross plots of CO_2 forcing and ΔMAT for our high-resolution $\delta^{11}B$–CO_2 record (**d**), published alkenone CO_2 data (**e**) and published low-resolution $\delta^{11}B$–CO_2 data (**f**). In each panel the slopes of regression lines fitted through the data are labelled (± 1 standard error, se). In **d** ice-core CO_2 data are shown as red open circles and Pliocene $\delta^{11}B$–CO_2 as open blue circles. In **e** and **f**, ice-core CO_2 data are shown in grey for clarity. In **e**, alkenone CO_2 data are from the following sources: ODP 1208 (orange[16]), ODP 806 (purple[16]); ODP 925 (brown[49]); ODP 999 (green circles[25]; green squares[26]). In **c** $\delta^{11}B$–CO_2 data are from ODP 999 (blue[25] and red[23]).
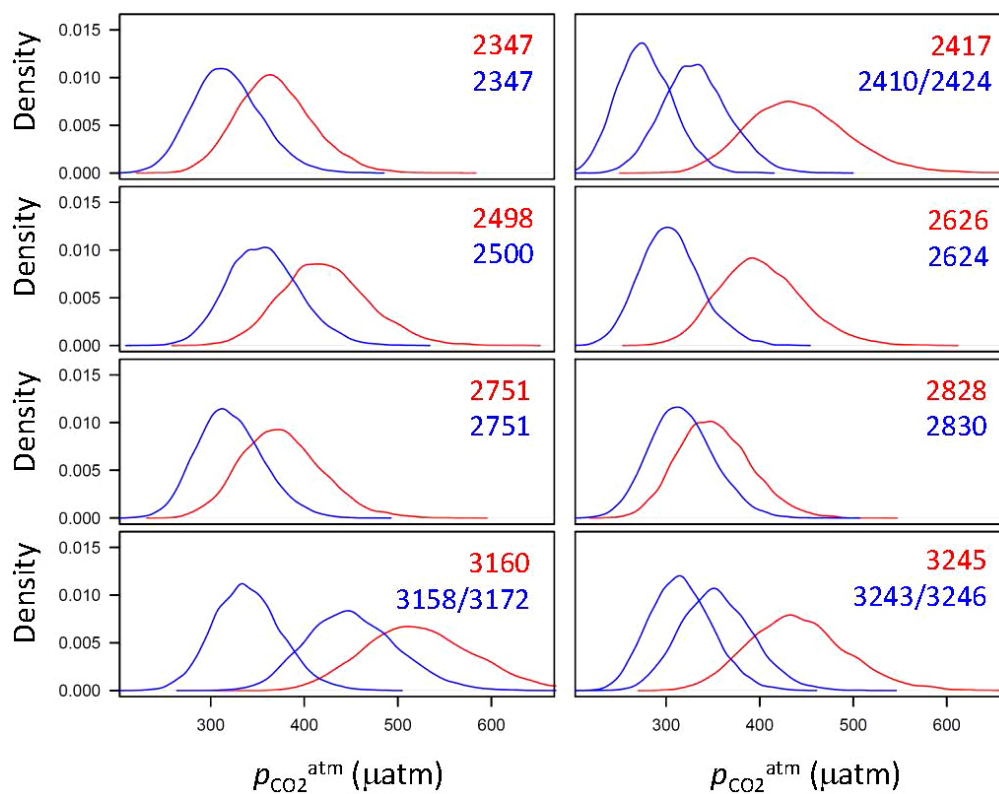
Extended Data Figure 3 | Probability density functions for equivalently aged samples from ODP Site 662 and ODP Site 999. Each panel, labelled with age (in units of kyr ago), shows the probability density function for a given estimate of $p_{CO_2}^{atm}$ from ODP Site 662 (red) and ODP Site 999 (blue). In most instances equal age samples are compared, but in some cases either where variability is high and/or equivalent age samples are absent, we show neighbouring samples from ODP Site 999 (for example, bottom left and right). This comparison indicates that although the mean $p_{CO_2}^{atm}$ of ODP 662 tends to be higher than ODP 999, there is always a high degree of overlap between the estimates from the two sites.
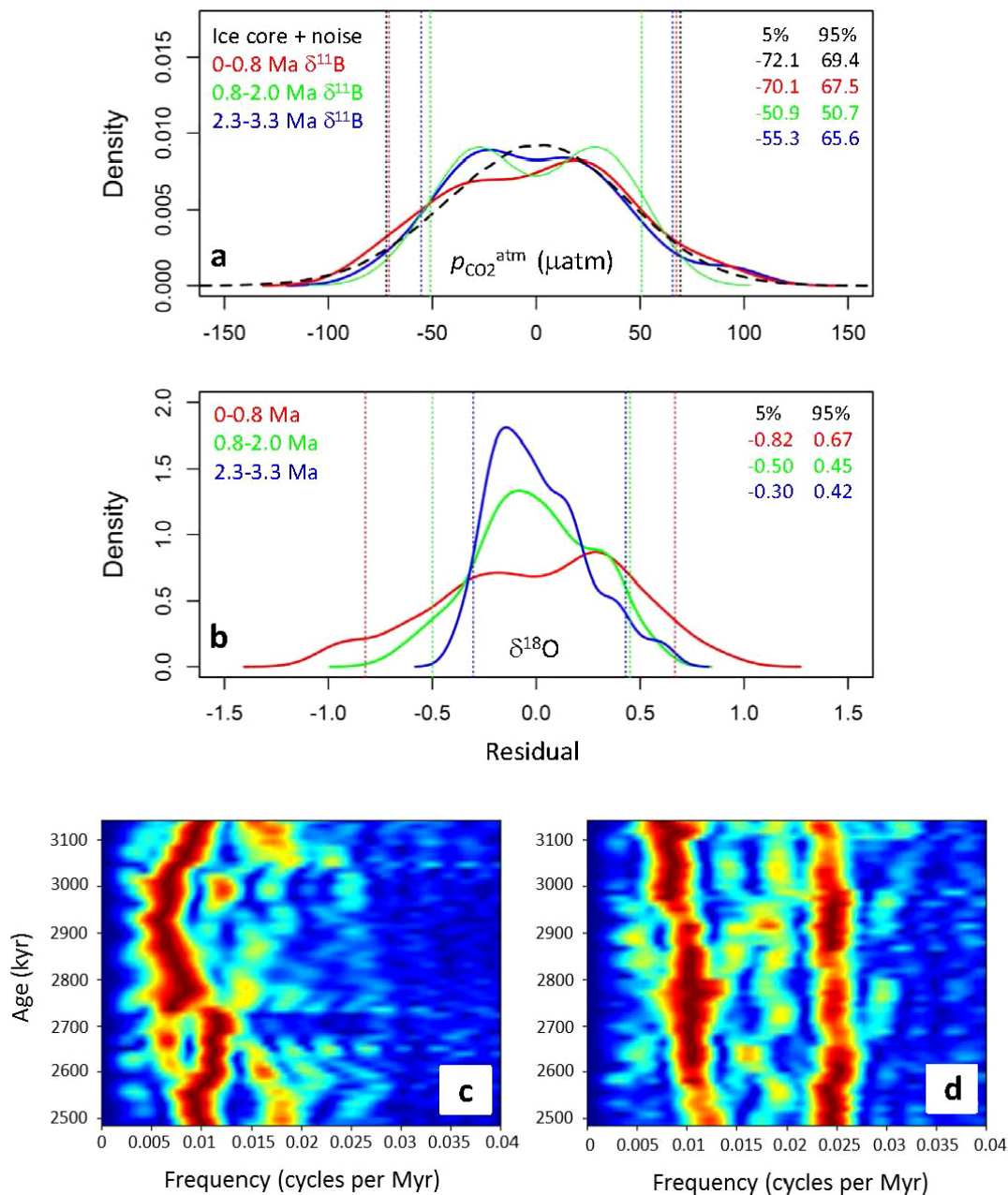
Extended Data Figure 4 | Probability density functions of $p_{CO_2}^{atm}$ and benthic $\delta^{18}O$ and time series analysis. **a**, Probability density functions of the residuals of $\delta^{11}B-p_{CO_2}^{atm}$ about the long-term trend for the late Pliocene (this study; blue line), the mid-Pleistocene[30] (green line) and late Pleistocene[19,20] (red line). Dashed vertical lines show the upper and lower limit (labelled) encompassing 90% of the data. The residual of the ice-core CO_2 record[27–29] about the long-term mean for 0–0.8 Myr ago plus a random noise equivalent to $\pm35\,\mu atm$ (the typical $\delta^{11}B-CO_2$ uncertainty) is shown as a black dashed probability density function. **b**, Probability density functions of the residual of LR04 benthic $\delta^{18}O$ from the long-term trend for the late Pleistocene (red),

mid-Pleistocene (green) and late Pliocene (blue). Dashed vertical lines show the upper and lower limit (labelled) encompassing 90% of the data. In contrast to $p_{CO_2}^{atm}$, $\delta^{18}O$ clearly exhibits an increase in variability over the last 3.3 Myr. **c**, **d**, Evolutive power spectral analyses of Pliocene $p_{CO_2}^{atm}$ (**c**) and resampled $\delta^{18}O$ (ref. 22) (**d**). The evolutive power spectra was computed using the fast Fourier transform of overlapping segments with a 300-kyr moving window. Before spectral analysis, all series were notch-filtered to remove the long-term trend (bandwidth = 0.005), and interpolated to 12-kyr intervals (the real resolution of our record is ~13.5 kyr).
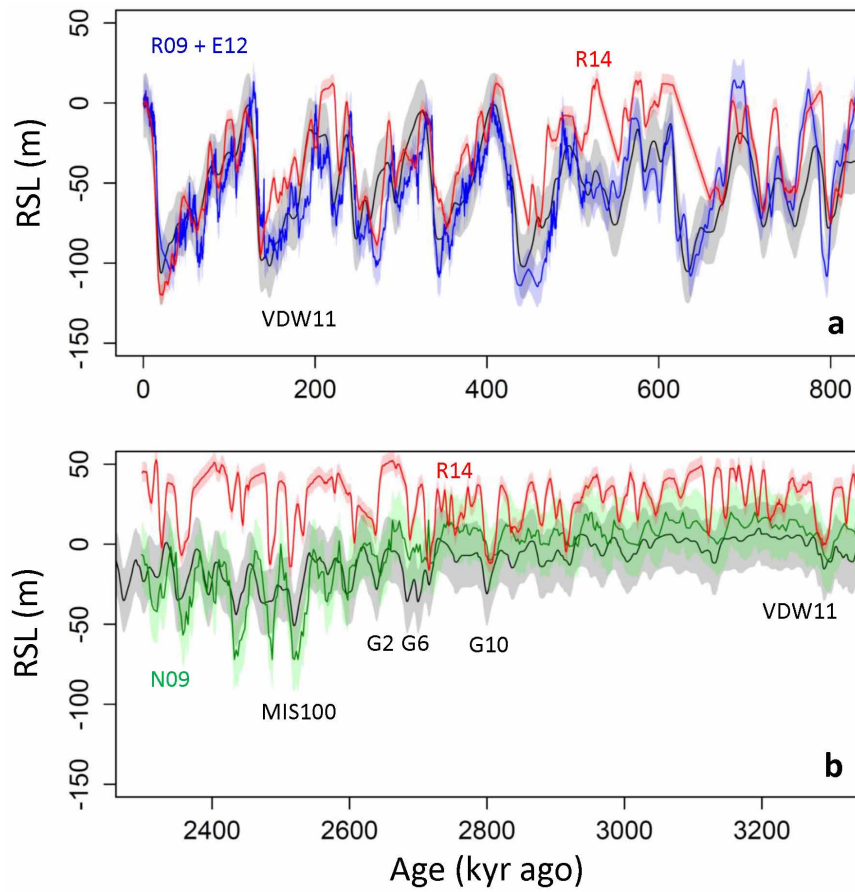
Extended Data Figure 5 | Summary of sea-level records used to calculate ΔF_LI. In **a** and **b** the red curve is from ref. 13 (R14) based on the planktic δ¹⁸O from the Mediterranean Sea and the methods developed for the Red Sea by ref. 93. We have removed those intervals identified as possible sapropel (organic-matter-rich sediments) events and linearly interpolated across gaps in the original record. The black curve is the sea-level record from an inversion of the benthic oxygen isotope record of ref. 76 (tuned to LR04 here) using an ice sheet model[35] (VDW11). The blue curve in **a** is based on the planktic/bulk δ¹⁸O from the Red Sea[44] for the interval 0–520 kyr and the paired Mg/Ca and benthic δ¹⁸O from the deep South Pacific for the interval 520–800 kyr (ref. 45) (R09+E12). The green curve in **b** is based on a scaling of the LR04 δ¹⁸O stack to indicators of sea level from sequence stratigraphy (ref. 46 recalculated by ref. 12). In each the uncertainty in the reconstruction at 95% confidence is shown by an appropriately coloured error band. Marine isotope stages mentioned in text are labelled. RSL, relative sea-level change (in metres), relative to the modern value.
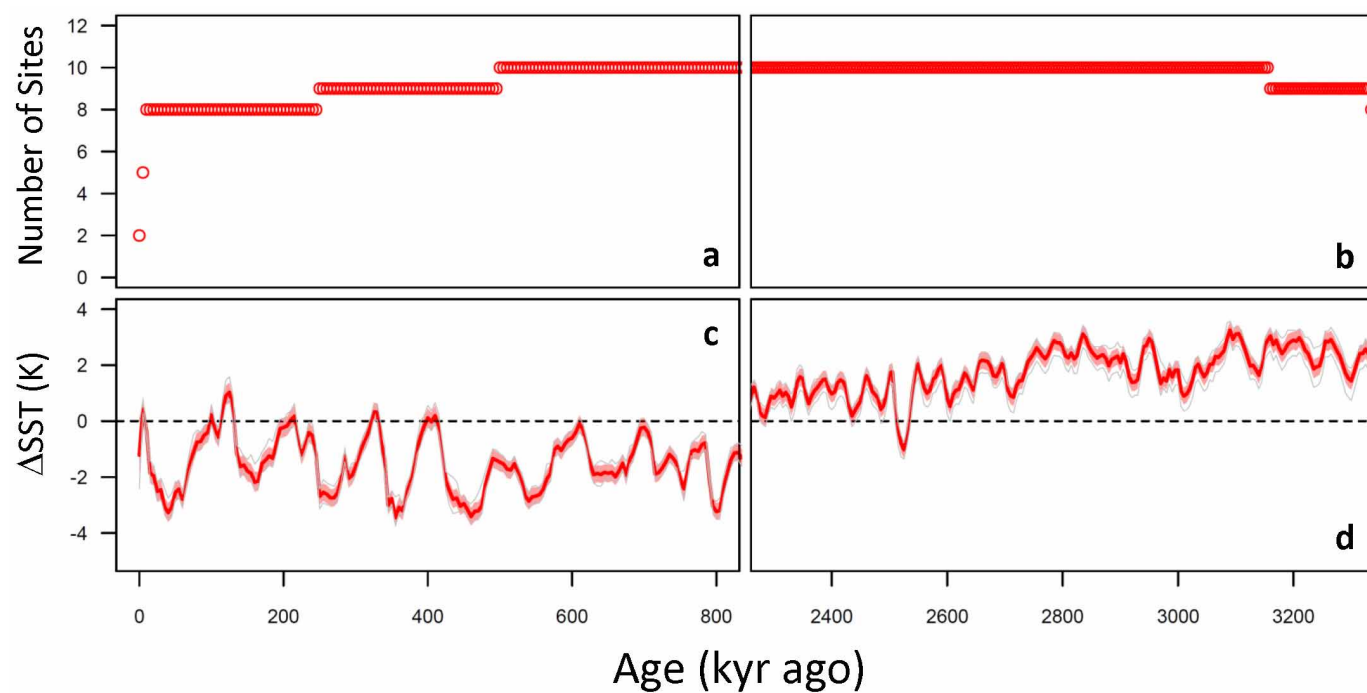
Extended Data Figure 6 | Stacked sea surface temperature record.
a, b, Number of records that contribute to the SST stack through time. c, d, Uncertainty in the SST stack due to analytical uncertainty (at 95% confidence; red band) and showing the influence of jacknifing (that is, removing one record at a time; grey lines show maximum and minimum). Note that the jacknifing illustrates that no single record has an undue influence on the SST stack.
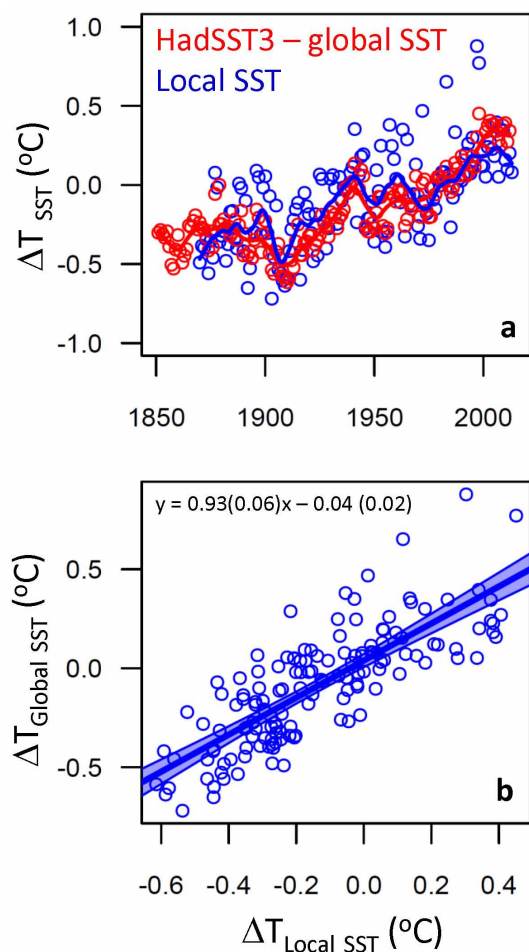
Extended Data Figure 7 | Comparison of global SST from the HadSST3 data set with SST HadISST1 from ODP sites. **a**, Historic global mean annual sea surface temperature anomaly from the HadSST3 data set[86,87] (red circles) and mean SST at locations above the ODP sites that make up the SST stack from HadISST1 (blue; local SST). Thick red and blue lines are non-parametric smoothers through both data sets. **b**, Cross plot of global mean annual SST and local SST. The regression line determined using linear regression has a slope of ~1 and intercept of close to 0, so local SST captures the global trend well. The shaded blue band in **b** represents the 95% confidence interval of the regression line.
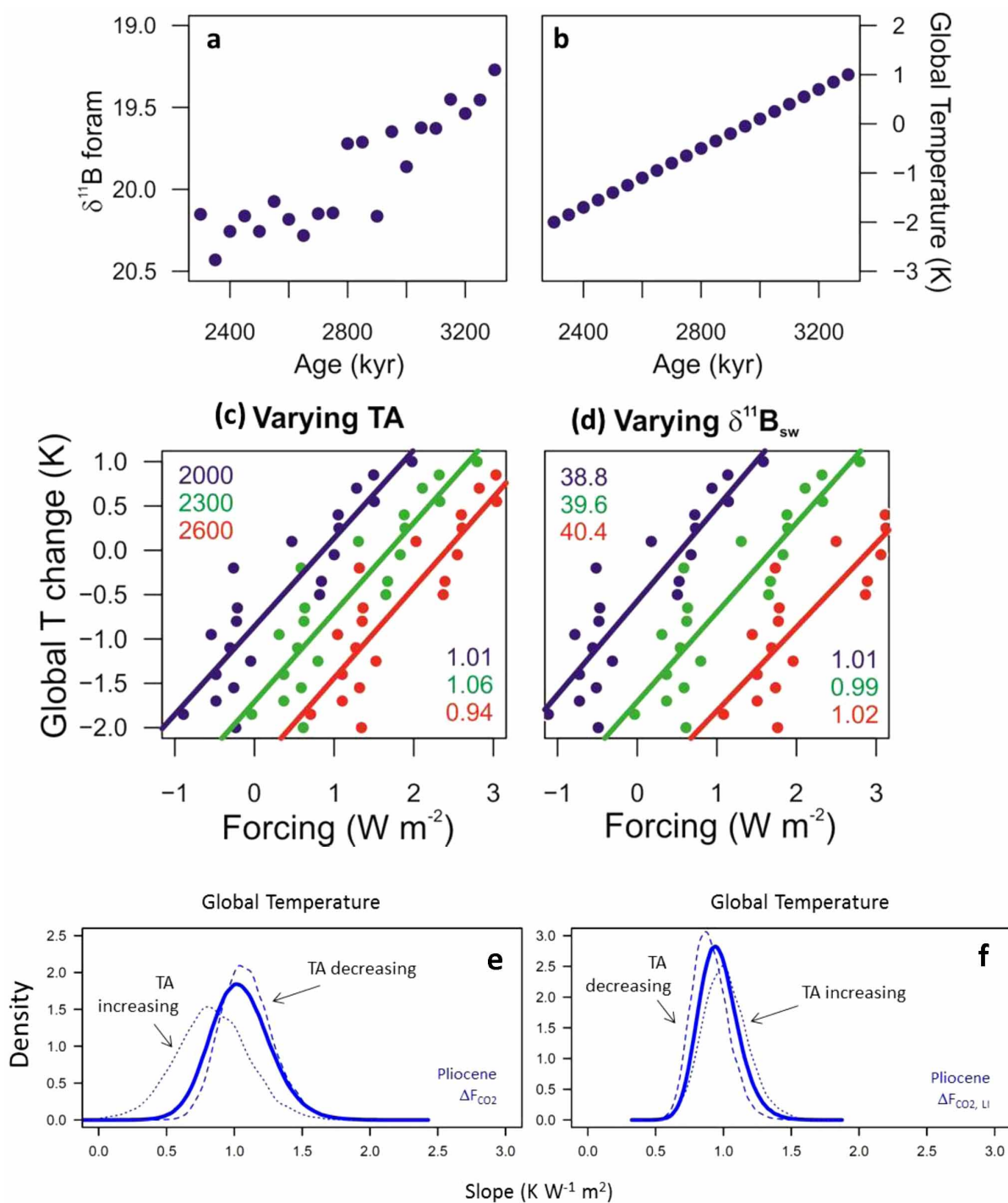
Extended Data Figure 8 | The influence of TA and $\delta^{11}B_{sw}$ on determinations of S^P using linear regression. **a**, **b**, Artificial $\delta^{11}B$ record (where $\delta^{11}B$ foram is the boron isotopic composition of an artificial foraminifera; **a**) and temperature record (**b**). **c**, **d**, Cross plot and regressions of $\delta^{11}B$–ΔF_{CO2} and global temperature for TA dramatically varying in the range 2,000–2,600 µmol kg^{-1} (TA; **c**) and $\delta^{11}B_{sw}$ from 38.8‰ to 40.4‰ (**d**). The slopes of the regressions, which are very similar regardless of parameter choice, are colour-coded and listed in the bottom right-hand corner of **c** and **d**. **e**, **f**, Probability density function of slope for regressions of Pliocene-aged ΔMAT against ΔF_{CO2} (**e**) and $\Delta F_{CO2,LI}$ (**f**), where TA is decreasing by 200 µmol kg^{-1} (dashed) and increasing by 200 µmol kg^{-1} (dotted). Note that despite large variations in TA the slope of the regressions do not change greatly.

ARTICLE

Biocontainment of genetically modified organisms by synthetic protein design

Daniel J. Mandell[1]*, Marc J. Lajoie[1,2]*, Michael T. Mee[1,3], Ryo Takeuchi[4], Gleb Kuznetsov[1], Julie E. Norville[1], Christopher J. Gregg[1], Barry L. Stoddard[4] & George M. Church[1,5]

Genetically modified organisms (GMOs) are increasingly deployed at large scales and in open environments. Genetic biocontainment strategies are needed to prevent unintended proliferation of GMOs in natural ecosystems. Existing biocontainment methods are insufficient because they impose evolutionary pressure on the organism to eject the safeguard by spontaneous mutagenesis or horizontal gene transfer, or because they can be circumvented by environmentally available compounds. Here we computationally redesign essential enzymes in the first organism possessing an altered genetic code (*Escherichia coli* strain C321.ΔA) to confer metabolic dependence on non-standard amino acids for survival. The resulting GMOs cannot metabolically bypass their biocontainment mechanisms using known environmental compounds, and they exhibit unprecedented resistance to evolutionary escape through mutagenesis and horizontal gene transfer. This work provides a foundation for safer GMOs that are isolated from natural ecosystems by a reliance on synthetic metabolites.

GMOs are rapidly being deployed for large-scale use in bioremediation, agriculture, bioenergy and therapeutics[1]. In order to protect natural ecosystems and address public concern it is critical that the scientific community implements robust biocontainment mechanisms to prevent unintended proliferation of GMOs. Current strategies rely on integrating toxin/antitoxin 'kill switches'[2,3], establishing auxotrophies for essential compounds[4], or both[5,6]. Toxin/antitoxin systems suffer from selective pressure to improve fitness through deactivation of the toxic product[7,8], while metabolic auxotrophies can be circumvented by scavenging essential metabolites from nearby decayed cells or cross-feeding from established ecological niches. Effective biocontainment strategies must protect against three possible escape mechanisms: mutagenic drift, environmental supplementation and horizontal gene transfer (HGT). Here we introduce 'synthetic auxotrophy' for non-natural compounds as a means to biological containment that is robust against all three mechanisms. Using the first genomically recoded organism (GRO)[9] we assigned the UAG stop codon to incorporate a non-standard amino acid (NSAA) and computationally redesigned the cores of essential enzymes to require the NSAA for proper translation, folding and function. X-ray crystallography of a redesigned enzyme shows atomic-level agreement with the predicted structure. Combining multiple redesigned enzymes resulted in GROs that exhibit markedly reduced escape frequencies and readily succumb to competition by unmodified organisms in non-permissive conditions. Whole-genome sequencing of viable escapees revealed escape mutations in a redesigned enzyme and also disruption of cellular protein degradation machinery. Accordingly, reducing the activity of the NSAA aminoacyl-tRNA synthetase in non-permissive conditions produced double- and triple-enzyme synthetic auxotrophs with undetectable escape when monitored for 14 days (detection limit 2.2×10^{-12} escapees per colony forming unit (c.f.u.)). We additionally show that while bacterial lysate supports the growth of common metabolic auxotrophs, the environmental absence of NSAAs prevents such natural products from sustaining synthetic auxotrophs. Furthermore, distributing redesigned enzymes throughout the genome reduces susceptibility

to HGT. When our GROs incorporate sufficient foreign DNA to overwrite the NSAA-dependent enzymes, they also revert UAG function, thereby preserving biocontainment by deactivating recoded genes. The general strategy developed here provides a critical advance in biocontainment as GMOs are considered for broader deployment in open environments.

Computational design of synthetic auxotrophs

We focused on the NSAA L-4,4′-biphenylalanine (bipA), which has a size and geometry unlike any standard amino acid, and a hydrophobic chemistry expected to be compatible with protein cores. We introduced a plasmid containing a codon-optimized version of the bipA aminoacyl-tRNA synthetase (*bipARS*)/tRNA$_{bipA}$ system[10] into a GRO (genomically recoded *E. coli* strain C321.ΔA (ref. 9)), thereby assigning UAG as a dedicated codon for bipA incorporation. Using a model of bipA in the Rosetta software for macromolecular modelling[11] we applied our computational second-site suppressor design protocol to 13,564 core positions in 112 essential proteins[12] with X-ray structures (Methods). We refined designs for cores that tightly pack bipA while maximizing neighbouring compensatory mutations predicted to destabilize the proteins in the presence of standard amino acid suppressors at UAG positions (Fig. 1a). We further required that candidate enzymes produce products that cannot be supplemented by environmentally available compounds. For example, we rejected *glmS* designs because glucosamine supplementation rescues growth of *glmS* mutants[13]. We selected designs of six essential genes for experimental characterization: adenylate kinase (*adk*), alanyl-tRNA synthetase (*alaS*), DNA polymerase III subunit delta (*holB*), methionyl-tRNA synthetase (*metG*), phosphoglycerate kinase (*pgk*) and tyrosyl-tRNA synthetase (*tyrS*). For all cases we designed oligonucleotides (Supplementary Table 1) encoding small libraries suggested by the computational models (Supplementary Table 2) and used them to directly edit the target essential gene in C321.ΔA using co-selection multiplex automated genome engineering (CoS–MAGE)[14]. Since *tyrS* featured the greatest number of compensatory mutations, we additionally

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Program in Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA. [4]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [5]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA.
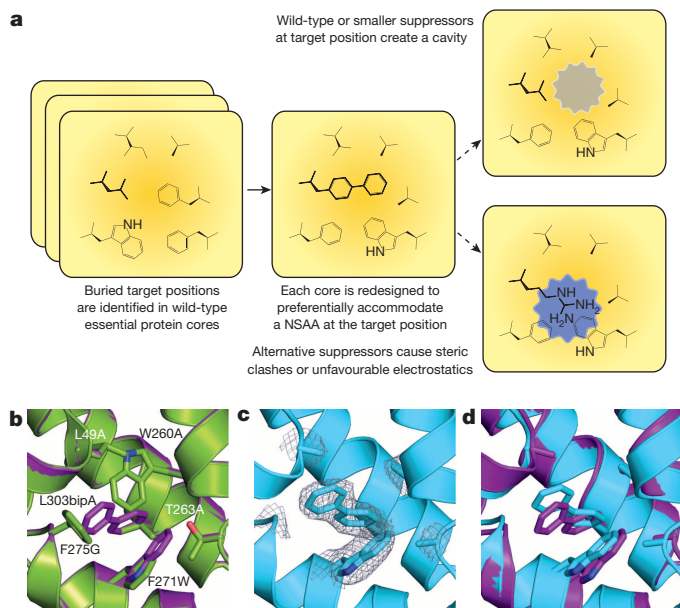*These authors contributed equally to this work.

Figure 1 | Computational design of NSAA-dependent essential proteins.
a, Overview of the computational second-site suppressor strategy.
b, Computational design of a NSAA-dependent tyrosyl-tRNA synthetase (purple) overlaid on the wild-type structure (green; PDB code 2YXN). Six substituted residues are shown in stick representation. **c**, X-ray crystallography of the redesigned synthetase with an electron density map ($2F_o − F_c$ contoured at 1.0σ) for substituted residues; substitution F236A is on a disordered loop and is not observed. **d**, The crystal structure of the redesigned enzyme (cyan) superimposed onto the computationally predicted model (purple).

synthesized eight computational *tyrS* designs and used them to replace the endogenous *tyrS* gene (Supplementary Table 3). We screened our CoS–MAGE populations for bipA-dependent clones by replica plating from permissive media (containing bipA and arabinose for *bipARS* induction) to non-permissive media (lacking bipA and arabinose) and validated candidates by monitoring kinetic growth in the presence and absence of bipA (Methods and Extended Data Fig. 1). Mass spectrometry confirmed the specific incorporation of bipA in redesigned enzymes (Methods and Extended Data Fig. 2). X-ray crystallography of a redesigned enzyme at 2.65 Å resolution (Protein Data Bank (PDB)[15] code 4OUD, Extended Data Table 1) shows atomic-level agreement with computational predictions (Fig. 1b–d, Extended Data Fig. 3 and Supplementary Discussion). Selectivity for bipA in a redesigned core was further confirmed by measuring soluble protein content when bipA is mutated to leucine (wild-type residue) or tryptophan (most similar natural residue to bipA by mass) (Methods and Extended Data Fig. 4).

Characterization of synthetic auxotrophs

We characterized the escape frequencies of eight strains by plating on non-permissive media with and without *bipARS* inducer arabinose (Fig. 2, Supplementary Tables 4, 5 and Methods). Escapees exhibiting varying fitness were detected by the emergence of colonies in the absence of bipA. Two *tyrS* variants (tyrS.d6 and tyrS.d7) and two *adk* variants (adk.d4 and adk.d6) showed robust growth in permissive conditions and low escape frequencies in the absence of bipA. Strain alaS.d5 showed only minor impairment in the absence of bipA, suggesting that near-cognate suppression of the UAG codon by endogenous tRNA or mis-charging of natural amino acids by BipARS is adequate to support growth. Consistent with this hypothesis, inserting a UAG immediately after the start codon (strain alaS.d5.startUAG) further impairs growth in the absence of bipA, although bipA dependence is readily overcome by mutational escape. *HolB* recombinants presented only the designed bipA mutation (holB.d1) and none of the compensatory mutations, suggesting that the intended compensation may be too destabilizing, or that the native amino acids at those positions may be critical for
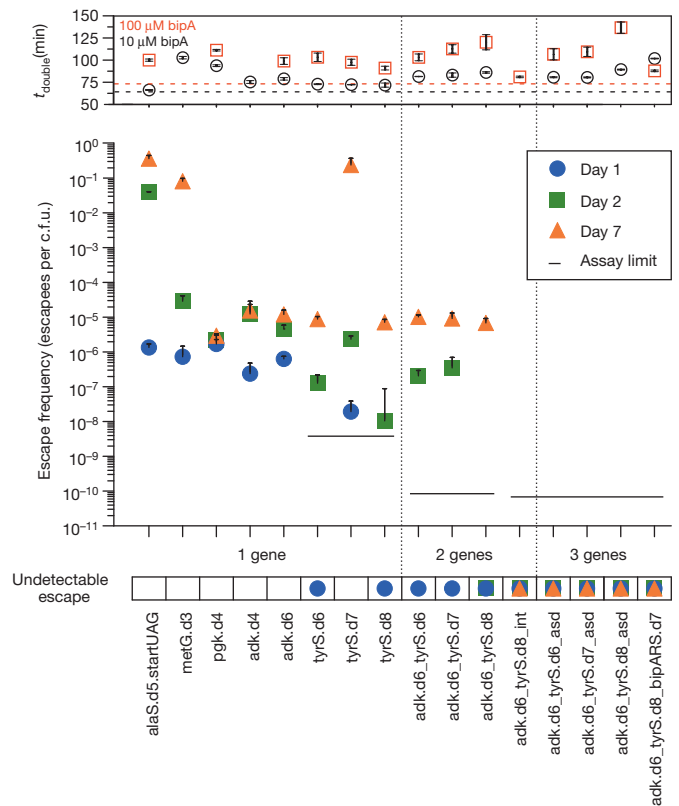
Figure 2 | Escape frequencies and doubling times of auxotrophic strains.
Escape frequencies are shown for engineered auxotrophic strains calculated as colonies observed per c.f.u. plated over three technical replicates on solid media lacking arabinose and bipA. Assay limit is calculated as 1/(total c.f.u. plated) for the most conservative detection limit of a cohort, with a single-enzyme auxotroph limit of 3.5×10^{-9} escapees per c.f.u., a double-enzyme auxotroph limit of 8.3×10^{-11} escapees per c.f.u. and a triple-enzyme auxotroph limit of 6.41×10^{-11} escapees per c.f.u. Positive error bars represent the s.e.m. of the escape frequency over three technical replicates (Methods). The top panel presents the doubling times for each strain in the presence of 10 μM or 100 μM bipA, with the parental strain doubling times represented by the dashed horizontal lines. No marker indicates undetectable growth. Positive and negative error bars represent the s.e.m.

function. The lack of compensation for bipA results in a strong and continuous selective pressure to incorporate standard amino acids at the bipA position, so holB.d1 was not carried forward.

We hypothesized that since the designed proteins have structurally distinct cores, each variant may favour different standard amino acids at the bipA position. Therefore, viable UAG suppressors for one enzyme may be deleterious for another. We sought to determine the distribution of standard amino acids accommodated at the UAG position in each variant to identify combinations of redesigned enzymes that could drive escape frequencies even lower. We cultured the top seven dependent strains in permissive media and used MAGE[16] to introduce all 64 codons at the UAG positions (Methods). For tyrS.d7 we collaterally introduced the V307A mutation observed in tyrS.d6, since the same oligonucleotide containing V307A was used to encode degeneracy for both strains at the UAG position, producing the eighth strain tyrS.d8. Immediately following electroporation, cells were shifted to non-permissive media so that recombinants with canonical amino acids replacing bipA would overtake the population according to their relative fitness.

We sampled the eight populations at 1-h and 4-h time points, at confluence, and at two subsequent passages to confluence (100-fold dilution in each passage), by which point the preferred genotypes emerged. Using next-generation sequencing we determined the relative abundance of all standard amino acid codons at the UAG positions for each time point (Extended Data Fig. 5 and Supplementary Table 6) and computed

the 'flatness' (Shannon entropy[17]) of each amino acid frequency distribution (Fig. 3). The two strains showing the greatest escape frequencies, alaS.d5 and metG.d3, also have the flattest amino acid frequency distributions. Correspondingly, the strains with the lowest escape frequencies exhibit peaked amino acid frequency distributions. These amino acid preference profiles show a strong relationship between structural selectivity for bipA and escape frequency, supporting the rationale underlying our computational strategy. Furthermore, they confirm our hypothesis that different redesigned protein cores will favour different standard amino acids. In particular, phenylalanine and tryptophan (aromatics) dominate tyrS.d7 and tyrS.d8 populations, whereas the other recombinants tend towards valine, leucine, isoleucine and methionine (aliphatics) (Fig. 3a). In agreement with these observations, we were able to isolate viable recombinants of adk.d6 containing leucine but not tryptophan at the bipA position, while also isolating viable recombinants of tyrS.d8 containing tryptophan but not leucine at the bipA position (Supplementary Table 7). In considering candidates for combination, we omitted alaS and metG due to their susceptibility for near-cognate suppression. We also determined that pgk mutants can grow robustly in the presence of pyruvate and/or succinate (Extended Data Fig. 6) even though they do not grow in lysogeny broth Lennox (LB^L)[12]. Since these carbon sources are environmentally available, pgk violates our definition of essentiality and we removed pgk.d4 from consideration. Finally, we excluded adk.d4 due to its poor survival at stationary phase (Supplementary Table 8, Supplementary Discussion). We therefore focused on combinations of tyrS.d6, tyrS.d7 and tyrS.d8 with adk.d6, all of which maintain robust growth in permissive media, show strong dependence for bipA, and are metabolically isolated from environmental compounds.

Combining tyrS designs with adk.d6 yielded three strains with no detectable escapees after 24 h, including adk.d6_tyrS.d8, which has undetectable growth after >72 h (detection limit 7.44×10^{-11} escapees per c.f.u., Fig. 2 and Supplementary Table 4). Colonies bearing the adk.d6_tyrS.d8 genotype were observed between 4 and 7 days of incubation, but showed severely impaired fitness when grown in non-permissive liquid culture and were readily outcompeted by prototrophic E. coli
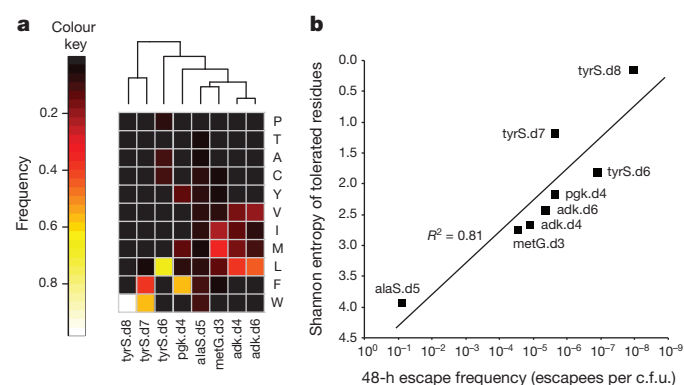
(Fig. 4 and Supplementary Discussion). The relative reductions in escape frequencies support the hypothesis that combining variants with distinct amino acid preferences at the UAG position decreases fitness of escapees. Even though tyrS.d6 and tyrS.d8 exhibit similar escape frequencies as single-enzyme auxotrophs, strain adk.d6_tyrS.d6 produces faster growing escapees (adk.d6 and tyrS.d6 share a preference for leucine) than strain adk.d6_tyrS.d8 (tyrS.d8 prefers tryptophan). Although tyrS.d7 and tyrS.d8 both prefer aromatic residues, tyrS.d7 exhibits a broader amino acid preference profile (Fig. 3a) and produces faster-growing escapees than tyrS.d8 (Fig. 2). Accordingly, adk.d6_tyrS.d8 yields the lowest escape frequency of the combined tyrS and adk variants.

Prevention of mutagenic escape

The appearance of escapee colonies from adk.d6_tyrS.d8 after >72 h suggests the emergence of rare genotypes conferring weak viability (doubling time ≥348 min) in the absence of bipA. To uncover mutagenic routes to escape we performed whole-genome sequencing on escapees of adk.d6, tyrS.d8 and adk.d6_tyrS.d8 (Methods, four escapees for each single-enzyme auxotroph and three escapees for the double-enzyme auxotroph were sequenced). We observed no mutations in the ribosome or tRNAs that could account for UAG translation in the absence of bipA, nor did we observe mutations to any designed amino acid positions. However, we identified a point mutation (A70V) to tyrS in all four tyrS.d8 escapees sequenced (Supplementary Table 9). The A70V mutation may improve packing of the tyrS.d8 catalytic domain in the context of a destabilized neighbouring helical bundle lacking bipA (Extended Data Fig. 7a). To validate this escape mechanism we produced strain tyrS.d8.A70V and performed an escape assay on non-permissive media. Within 5 days of plating, we observed colony formation from all plated cells (Extended Data Fig. 7b), confirming that A70V is an escape mechanism for tyrS.d8. The A70V mutant of tyrS.d8 does not impair fitness in permissive conditions (Supplementary Table 10), so the genotype spontaneously arises as a neutral mutation within the fitness landscape by replication errors. However, targeted sequencing of the tyrS gene in eight additional tyrS.d8 escapees did not reveal the A70V mutation, suggesting that A70V is not the only escape mechanism for tyrS.d8.

Whole-genome sequencing of adk.d6 and adk.d6_tyrS.d8 escapees revealed disruptive mutations to Lon protease in all seven cases. One clone contained a frame shift and another contained a non-synonymous substitution (L611P) within the lon gene. The remaining five cases had a transposable element inserted within the promoter of lon. Targeted

Figure 3 | Structural specificity at designed UAG positions in eight NSAA-dependent enzymes correlates with escape frequencies. **a**, Amino acid preferences at UAG positions in eight synthetic auxotrophs were determined by replacing the UAG codon with full NNN degeneracy and then sequencing the resulting populations with an Illumina MiSeq. Frequencies of each amino acid as a fraction of total sequences observed after three 1:100 passages to confluence are shown (top 11 most frequent amino acids only). Samples are clustered by Euclidean distance between amino acid frequencies. The frequency of an amino acid reports on the fitness conferred by the corresponding natural amino acid suppressor at the UAG position relative to all other amino acids. **b**, Shannon entropy was computed over the distributions of amino acids preferred at the UAG positions of the eight single-enzyme auxotrophs and plotted against the 48-h escape frequency for each strain. Entropy correlates log-linearly with escape frequency, suggesting that enzyme cores with high structural specificity for bipA at the UAG position have fewer fit evolutionary routes to escape. Strains alaS.d5 and metG.d3 have a deactivated mutS gene.

Figure 4 | Competition between synthetic auxotroph escapees and prototrophic E. coli. C321.ΔA was competed in the absence of bipA against escapees from a single-enzyme bipA auxotroph (pgk.d4, moderate NSAA dependence), or from a double-enzyme bipA auxotroph (adk.d6_tyrS.d8, strong bipA-dependence). Populations were seeded with 100-fold excess escapees and grown for 8 h in non-permissive conditions. The populations were evaluated using flow cytometry for episomally expressed fluorescent proteins at t = 0 and t = 8 h. Results from separate competition experiments against three different escapees are shown for each synthetic auxotroph. **a**, pgk.d4 escapees continue to expand in a mixed population with C321.ΔA after 8 h. **b**, adk.d6_tyrS.d8 escapees are rapidly outcompeted by C321.ΔA, which overtakes the population after 8 h.

sequencing characterized the insertion sequence in at least one clone as IS186, exactly recapitulating the Lon protease deficiency of *E. coli* BL21[18]. We validated Lon disruption as an escape mechanism using λ Red-mediated recombination to replace *lon* with a kanamycin resistance gene (*kan^R*) in adk.d6, tyrS.d8 and adk.d6_tyrS.d8. Recombinants were replica plated from permissive to non-permissive media containing kanamycin. Colony PCR confirmed that 27 of 27 non-bipA-dependent colonies screened (9 escapees per dependent strain) had Lon deleted by *kan^R*.

Since the Lon protease is the primary apparatus for bulk degradation of misfolded proteins in the *E. coli* cytoplasm[19], we hypothesized that its disruption would allow the persistence of poorly folded adk.d6 and tyrS.d8 proteins when standard amino acids are incorporated in place of bipA. We further hypothesized that basal UAG suppression from the promiscuous activity of pEVOL-BipARS produced sufficient full-length protein to support viability in the absence of Lon-mediated degradation. To test this hypothesis and safeguard against Lon-mediated escape we pursued two independent strategies to reduce the activity of BipARS in non-permissive conditions. First, we reduced the gene copy number approximately tenfold by moving *bipARS* and tRNA$_{bipA}$ from the p15A pEVOL plasmid to the genome of adk.d6, producing the strain adk.d6_int (Methods). Second, we applied our computational second-site suppressor strategy to residue V291 in *bipARS* (homologous to L303bipA in our *tyrS* designs) and reintroduced it into adk.d6 on the pEVOL vector, producing strain adk.d6_bipARS.d7 (Methods). This latter strategy produced a BipARS variant that requires bipA for folding and function, thereby abrogating residual activity towards standard amino acids in the absence of bipA. Both strategies resulted in a >200-fold reduction in 7-day escape frequency (Supplementary Table 4). Introducing tyrS.d8 to these strains produced double- and triple-enzyme synthetic auxotrophs adk.d6_tyrS.d8_int and adk.d6_tyrS.d8_bipARS.d7 that exhibited undetectable escape when monitored for 14 days (Fig. 2 and Supplementary Table 4, detection limit 2.2×10^{-12} escapees per c.f.u.). Both strains also showed undetectable escape in the presence of arabinose (Supplementary Table 5), and presented no fitness impairment relative to the parental adk.d6_tyrS.d8 synthetic auxotroph (Supplementary Table 4, doubling times of 57 and 55 min).

Protection from natural supplementation

To compare synthetic auxotrophy to current biocontainment practices we generated natural metabolic auxotrophs by knocking out *asd* and *thyA* genes from an MG1655-derived *E. coli* strain (EcNR1). The *asd* knockout renders the strain dependent on diaminopimelic acid (DAP) for cell-wall biosynthesis[4], while the *thyA* knockout deprives the cell of thymine, an essential nucleobase[20]. These well studied auxotrophies are commonly incorporated into biocontainment strategies[4,6]. In agreement with previous studies[4,6], the *asd* knockout shows strong dependence on its requisite metabolite, with a 7-day escape frequency of 8.97×10^{-9} escapees per c.f.u. (Supplementary Table 11). Knocking out *thyA* from this strain to produce a double-enzyme auxotroph did not reduce the 7-day escape frequency (8.79×10^{-9} escapees per c.f.u.). Nevertheless, metabolic strategies could complement synthetic auxotrophies to improve escape frequencies in defined ecological niches. To test this principle we knocked out *asd* from the double-enzyme synthetic auxotrophs of *adk* and *tyrS* resulting in three triple-enzyme auxotrophs (adk.d6_tyrS.d6_asd, adk.d6_tyrS.d7_asd and adk.d6_tyrS.d8_asd) that grew robustly in permissive conditions but showed undetectable escape after 7 days on media lacking bipA and DAP (Fig. 2 and Supplementary Tables 4 and 5, detection limit 6.4×10^{-11} escapees per c.f.u.).

While bacterial growth assays are often carried out in variations of media enriched with yeast extract, GMOs are increasingly deployed among a diversity of ecosystems that may provide opportunities for scavenging or cross-feeding essential metabolites. To compare metabolic and synthetic auxotroph strategies in an environment mimicking endogenous bacterial communities we grew engineered variants of both natural and synthetic auxotrophs in LB^L containing *E. coli* lysate

(Methods). We hypothesized that since DAP is an essential component of the bacterial cell wall, the *Δasd* strains may scavenge sufficient DAP from *E. coli* lysate to complement the auxotrophy. As anticipated, metabolic auxotrophs obtained sufficient nutrients from the yeast/tryptone (LB^L) and the bacterial remnants (lysate) to support exponential growth (Extended Data Fig. 6e–h), while the synthetic auxotrophs failed to circumvent their dependencies. These results highlight the importance of establishing auxotrophies for compounds that are not environmentally available, and of ensuring the metabolic essentiality of enzymes intended to confer dependence.

Resistance to horizontal gene transfer

HGT is an important mechanism of evolution in any genetically rich environment[21]. We developed a conjugation escape assay to assess how DNA transfer within an ecosystem enables a GMO to escape biocontainment. Whereas any recombination event that replaces an inactivated gene could overcome metabolic auxotrophies[22], we hypothesized that conjugal escape would be disfavoured in GROs because donor DNA replacing bipA-dependent genes would also overwrite crucial genetic elements involved in genetic code reassignment (Fig. 5a). For example, reintroducing UAG stop codons into essential genes without restoring release-factor-1-mediated translational termination could be deleterious[9] or lethal[23]. Furthermore, reintroducing release factor 1 would result in competition between bipA incorporation and translational termination, undermining the recoded functions of the GRO.

To simulate a worst-case scenario in ecosystems containing a rich source of conjugal donors, we used Tn5 transposition to integrate an origin of transfer (*oriT*) into a population of *E. coli* MG1655 conjugal donor strains. We isolated a population of ~450 independent clones (one *oriT* for every ~10 kilobase portion of the 4.6 megabase (Mb) genome) and sequenced the flanking genomic regions of 96 donor colonies to confirm that *oriT* integration was well distributed throughout the population. We then conjugated this donor population into our auxotrophic strains at a ratio of 1 donor to 100 recipients to increase the probability that conjugal transfer will initiate from one *oriT* position per recipient. Conjugation was performed for durations of 50 min and 12 h (average conjugation times predicted to transfer 0.5 and 7.2 genomes) to simulate a single conjugal interaction and an ecological worst-case scenario, respectively. Conjugal escapees were selected on non-permissive media, and 23 alleles distributed throughout the genome (Fig. 5a) were screened using multiplex allele-specific colony PCR (mascPCR) to assess how much of the recoded genome is replaced by wild-type donor DNA.

Conjugal escape frequency decreases as the number of auxotrophic gene variants increases (Fig. 5b, top panel and Extended Data Fig. 8), consistent with larger portions of the genome that must be overwritten for conjugal escape of the multi-enzyme auxotrophs (Fig. 5b, bottom panel). The 12-h conjugations effect higher escape frequencies than do the 50-min conjugations, and the 12-h conjugations produce a larger diversity of conjugal escape genotypes, consistent with an increased opportunity to initiate new conjugal transfers during the mating period. Encouragingly, all 50-min conjugal escapees from multi-enzyme auxotrophs exhibit the wild-type donor sequence at all 23 assayed alleles (Fig. 5b, bottom panel and Supplementary Table 12), resulting in the reintroduction of release factor 1 and its UAG-mediated translational termination function. This collateral replacement of recombinant genomic DNA could be extended to other recombinant payloads such as toxins, antibiotic resistance genes, catabolic and genome editing enzymes, and orthogonal aminoacyl-tRNA synthetase/tRNA pairs used for NSAA incorporation.

Discussion

Effective biocontainment mechanisms for GMOs should place high barriers between modified organisms and the natural environment. Our NSAA design strategy produces organisms with an altered chemical language that isolates them from natural ecosystems. By conferring dependence on synthetic metabolites at the level of protein translation,
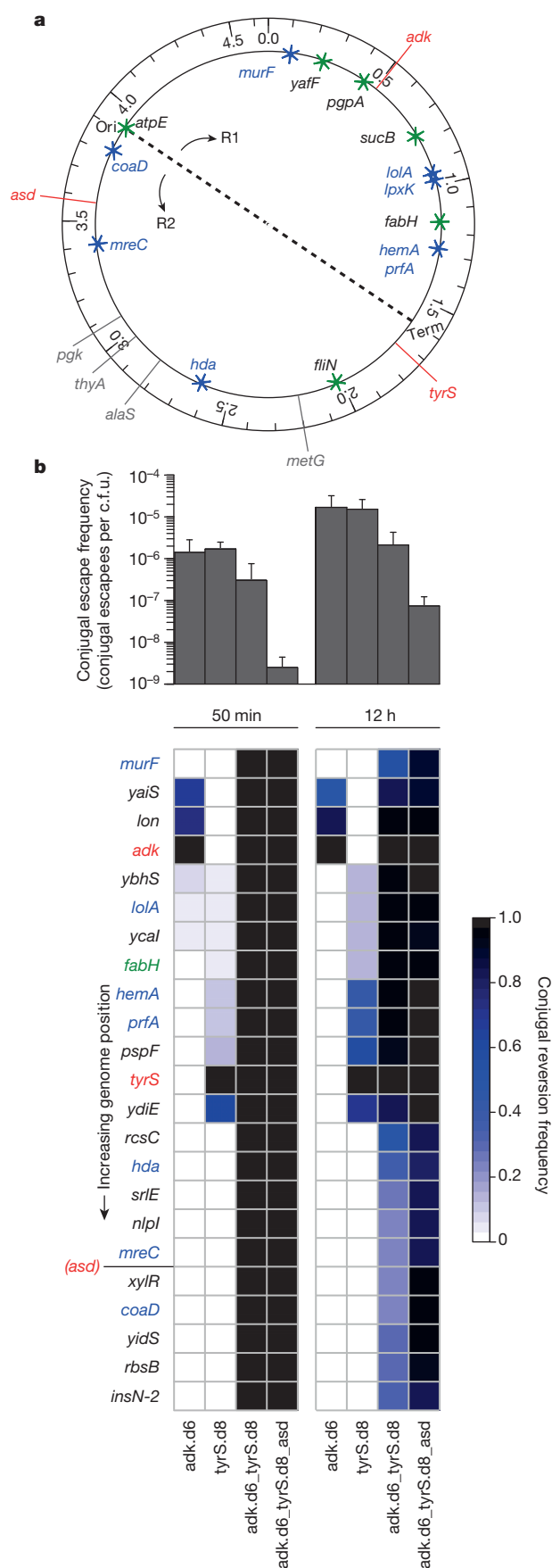
Figure 5 | Synthetic auxotrophy and genomic recoding reduce HGT-mediated escape. a, The positions of key alleles are plotted to scale on the genome schematic. Red lines indicate auxotrophies used in the multi-enzyme auxotrophs and grey lines indicate other auxotrophies that were not included in this assay. Asterisks indicate important alleles associated with the reassignment of UAG translation function (blue are essential genes and green are potentially important genes[9]). Conjugation-mediated reversion of the UAA codons back to the wild-type UAG is expected to be deleterious unless the natural UAG translational termination function is reverted. R1 and R2 denote replicores 1 and 2, respectively. **b,** Combining multiple synthetic auxotrophies in a single genome requires a large portion of the genome to be overwritten by wild-type donor DNA, reducing the frequency of conjugal escape (top panel) and increasing the likelihood of overwriting the portions of the genome (bottom panel) that provide expanded biological function (for example, *prfA* encodes release factor 1, which mediates translational termination at UAG codons). Positive error bars indicate standard deviation.

second-site mutations, escapees are rare and are unfit to out-compete prototrophic microbial communities. In part, robustness emerges from simplicity: our most escape-resistant synthetic auxotrophs contain only 32 (adk.d6_tyrS.d8_int) and 49 (adk.d6_tyrS.d8_bipARS.d7) base pair substitutions across the 4.6 Mb parental genome and *bipARS*, with no essential genes deleted or toxic products added. Furthermore, NSAA-based biocontainment with bipA only modestly increases the cost per litre of proliferating culture (Extended Data Table 2).

This work highlights the delicate balance required to engineer essential proteins that are conditionally stabilized by a single NSAA. The design must confer sufficient instability in non-permissive conditions to deactivate the protein, while providing functional stability in the presence of the correct NSAA. Future design strategies could include polar or charged NSAAs to engineer hydrogen bonds requiring exquisite geometric specificity[24] for folding, enzyme–substrate interactions, or macromolecular associations. This approach may reduce susceptibility to suppressors, although fewer protein microenvironments may accommodate the burial of charged or polar residues. Reassigning additional codons would permit the incorporation of multiple NSAAs that confer dependence either in different structural motifs or in participation of a joint chemistry. Eventually, organisms with orthogonal genomic chemistries including expanded genetic alphabets[25] and their associated replication machinery could provide additional layers of isolation[26].

Our results demonstrate that mutational escape frequency under laboratory growth conditions is a necessary but insufficient metric to evaluate biocontainment strategies. Many genes considered to be essential have functions that can be complemented by environmental compounds, as demonstrated here for auxotrophies of natural (*asd*, *thy*) and designed (*pgk.d4*) enzymes. Furthermore, localizing biocontainment mechanisms to a small portion of the genome increases susceptibility to escape by uptake of foreign DNA. Distributing multiple NSAA-dependent enzymes throughout a recoded genome acts as a genomic safeguard against escape by HGT, and demands that conjugal escape replaces large portions of the recipient genome. This collateral replacement of GMO genomic DNA could be exploited to delete recombinant payloads upon exposure to conjugal donors in the environment. Additionally, by recoding restricted payloads with essential UAG codons, they can be prevented from functioning in natural organisms. Therefore, the expanded genetic code of GROs can be exploited both to prevent their undesired survival in natural ecosystems and to block incoming and outgoing HGT with natural organisms.

folding and function, synthetic auxotrophy addresses the need for GMOs that are refractory to mutational escape, metabolic supplementation and HGT. Because our NSAAs are incorporated into essential enzymes with

1. Moe-Behrens, G. H., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4,** 5 (2013).
2. Molin, S. *et al.* Conditional suicide system for containment of bacteria and plasmids. *Nature Biotechnol.* **5,** 1315–1318 (1987).

3. Li, Q. & Wu, Y.-J. A fluorescent, genetically engineered microorganism that degrades organophosphates and commits suicide when required. *Appl. Microbiol. Biotechnol.* **82,** 749–756 (2009).
4. Curtiss, R., III. Biological containment and cloning vector transmissibility. *J. Infect. Dis.* **137,** 668–675 (1978).
5. Ronchel, M. C. & Ramos, J. L. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl. Environ. Microbiol.* **67,** 2649–2656 (2001).
6. Wright, O., Delmans, M., Stan, G. B. & Ellis, T. GeneGuard: a modular plasmid system designed for biosafety. *ACS Synth. Biol.* http://dx.doi.org/doi:10.1021/sb500234s (13 May 2014).
7. Knudsen, S. *et al.* Development and testing of improved suicide functions for biological containment of bacteria. *Appl. Environ. Microbiol.* **61,** 985–991 (1995).
8. Pasotti, L., Zucca, S., Lupotto, M., Cusella De Angelis, M. G. & Magni, P. Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *J. Biol. Eng.* **5,** 8 (2011).
9. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–360 (2013).
10. Xie, J., Liu, W. & Schultz, P. G. A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem.* **46,** 9239–9242 (2007).
11. Renfrew, P. D., Choi, E. J., Bonneau, R. & Kuhlman, B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS ONE* **7,** e32637 (2012).
12. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
13. Wu, H. C. & Wu, T. C. Isolation and characterization of a glucosamine-requiring mutant of *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. *J. Bacteriol.* **105,** 455–466 (1971).
14. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* (2012).
15. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
16. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
17. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379–423 (1948).
18. saiSree, L., Reddy, M. & Gowrishankar, J. IS*186* insertion at a hot spot in the *lon* promoter as a basis for *lon* protease deficiency of *Escherichia coli* B: identification of a consensus target sequence for IS*186* transposition. *J. Bacteriol.* **183,** 6943–6946 (2001).
19. Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol. Microbiol.* **40,** 397–413 (2001).
20. Steidler, L. *et al.* Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nature Biotechnol.* **21,** 785–789 (2003).
21. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244 (2011).
22. Wollman, E. L., Jacob, F. & Hayes, W. Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harb. Symp. Quant. Biol.* **21,** 141–162 (1956).
23. Mukai, T. *et al.* Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38,** 8188–8195 (2010).
24. Kortemme, T., Morozov, A. V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *J. Mol. Biol.* **326,** 1239–1259 (2003).
25. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509,** 385–388 (2014).
26. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered Biology. *FEBS Lett.* **586,** 2199–2206 (2012).

Author Contributions D.J.M., M.J.L., M.T.M. and G.M.C. conceived the project and designed the study, with D.J.M. as computational lead and M.J.L. as experimental lead. D.J.M. computationally designed synthetic auxotrophs, performed strain engineering, characterized escape frequencies and fitness of synthetic auxotrophs, performed western blot analyses and prepared samples for mass spectrometry and X-ray crystallography. M.J.L. performed strain engineering, performed site-saturation mutagenesis at UAG positions, performed whole-genome sequencing of escapees, validated escape mechanisms and assessed HGT by conjugation. M.T.M. measured escape frequencies and fitness of natural metabolic auxotrophs, performed competition assays and assessed HGT by conjugation. R.T. and B.L.S. crystallized tyrS.d7 and determined the X-ray structure. G.K. analysed whole-genome sequencing data of escapees. J.E.N. and C.J.G. developed the *tdk* selection protocol. D.J.M., M.J.L. and M.T.M. wrote the paper.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 4OUD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.M.C. (gchurch@genetics.med.harvard.edu).

METHODS

Essential protein selection and buried residue determination. Candidate genes were selected by searching the Keio collection[12] of comprehensive single-gene *E. coli* K-12 knockouts for genes classified as essential. X-ray structures were identified by mapping essential gene GenBank[27] protein genInfo identifiers (PIDs) to PDB[15] entries through the UniProtKB[28]. In cases of multiple PDB entries the highest-resolution structure was selected. 112 high-resolution X-ray structures (resolution ≤2.8 Å) were analysed. Structures were pre-processed to remove alternative side-chain conformers (the first listed conformer was kept), to remove atoms without occupancy, to remove heteroatoms, to convert selenomethionines to methionines, and to remove chains other than the first listed chain of the essential protein. The solvent-accessible surface area (SASA) of each position in each candidate structure was calculated using the PyRosetta[29] interface to the Rosetta SasaCalculator class with a 1.0 Å probe radius (a radius smaller than 1.4 Å allowed finer sampling of spaces around candidate positions). Positions were considered buried if their SASA was not more than 20% of the residue-specific average SASA value from a 30-member random ensemble of Gly-X-Gly peptides, where X is the residue type, as determined by the GETAREA method[30,31]. The average SASA values are Ala, 64.9; Arg, 195.5; Asn, 114.3; Asp, 113.0; Cys, 102.3; Gln, 143.7; Glu, 141.2; His, 154.6; Ile, 147.3; Gly, 87.2; Leu, 146.2; Lys, 164.5; Met, 158.3; Phe, 180.1; Pro, 105.2; Ser, 77.4; Thr, 106.2; Trp, 224.6; Tyr, 193.1; Val, 122.3. By these criteria 13,564 residues in the data set were considered buried.

Design and refinement of NSAA-dependent proteins. The side chains of each structure were relaxed into local minima of the Rosetta forcefield by the Rosetta sidechain_min application (Rosetta command lines, below). Three separate design simulations were then carried out for each target buried position using RosettaDesign[32]. The first simulation sets the target position to L-4,4′-biphenylalanine (bipA), as implemented in Rosetta (residue type B30)[11], and sets the surrounding residues to either redesign (varies both amino acid identity and side-chain conformation) or repack (varies only the conformation of the wild-type amino acid). Residues with Cα atoms within 6 Å of the target position, or with Cα atoms within 8 Å of the target position and Cβ atoms closer than the Cα atom to the target position, were set to redesign. Residues with Cα atoms within 10 Å of the target position, or with Cα atoms within 12 Å of the target position and Cβ atoms closer than the Cα atom to the target position, were set to repack. All other side chains were fixed at their minimized coordinates, together with all backbone atoms. The resulting energy terms were appended with the target position SASA as calculated by the PyRosetta SasaCalculator with a 1.0 Å probe radius. We term the Rosetta scores of these designs 'compensated scores'. In the second simulation, the same calculation is performed, except all positions previously set to redesign are restricted only to repack. We term the resulting scores 'uncompensated scores'. The difference between the 'compensated score' and the 'uncompensated score' reports on the extent to which the target site must change to accommodate bipA. In the third simulation, the target position maintains its wild-type identity, all coordinates are fixed at the positions output by the sidechain_min application, and the structure is rescored using the same scoring parameters as the other two simulations (Rosetta command lines, below). We term the resulting scores 'wild-type scores'. The difference between the compensated score and the wild-type score reports on the predicted stability of the redesigned core relative to the wild-type structure.

The design goal was to obtain variants that are functionally stable with bipA at the designed position, while being maximally destabilized with a natural amino acid at the bipA position. Accordingly, designs were filtered for the following criteria:

(1) The minimized wild-type score must be less than 10 Rosetta energy units to ensure that the starting structure can be reasonably modelled with the Rosetta forcefield.

(2) The compensated score must be less than or equal to the wild-type score, to select for redesigned cores that do not destabilize the protein relative to the wild-type sequence.

(3) The uncompensated score must be greater than the wild-type score, to ensure that compensatory mutations are necessary to accommodate bipA.

(4) The compensated score must be less than the uncompensated score, to select for compensatory mutations that improve the stability of the core in the presence of bipA relative to the uncompensated mutant. This requirement also selects for sequences that reduce the fitness of suppressors at the compensatory positions.

(5) The SASA score must be <0.75, to select for cores that tightly pack around bipA, both to select for stability in the presence of bipA and to reduce the fitness of standard amino acids at the bipA or compensatory positions.

The designs for 16 engineered UAG sites in 12 enzymes meeting these criteria were then ranked by the difference between compensated score and uncompensated score, as a key metric for bipA dependence, and were further filtered by the following criteria based on known structural and functional data from the literature:

(1) The redesigned residues must not participate in ligand binding, catalysis or be required for allosteric signal transduction via conformational rearrangements.

(2) The product of the reaction must not be environmentally available.

(3) The product of the reaction must not be complementable by another environmentally available molecule.

Using these criteria 10 designs were subject to refinement. Positions to design, repack, or revert to wild-type were selected by visual inspection. A second round of fixed backbone design was then applied to generate 100 designs from each unrefined structure (Rosetta command lines, below). Designs from six enzymes were carried forward for experimental characterization. Frequently occurring mutations in the refined designs assessed by visual inspection were included in MAGE oligonucleotides (Supplementary Table 1). For *tyrS*, eight additional all-atom designs were encoded by PCR primers for gene assembly (Supplementary Table 3).

Rosetta command lines. All Rosetta calculations were performed with Rosetta version 48561.

Example command line for preparative side-chain minimization of scaffold structures:

sidechain_min.linuxgccrelease -database __ROSETTA_DATABASE__ -loops:: input_pdb __PREFIX__.pdb -output_tag __PREFIX__ -ex1 -ex2 -overwrite

Example command line for wild-type score:

score.linuxgccrelease -database __ROSETTA_DATABASE__ -l *.pdb -score: hbond_His_Phil_fix -in:file:fullatom -no_optH -no_his_his_pairE -score:weights mm_std

Example command line for initial design for compensated score and uncompensated score:

fixbb.linuxgccrelease -database __ROSETTA_DATABASE__ -ex1 -ex2 -s __ PREFIX__.pdb -resfile __PREFIX__.resfile -minimize_sidechains -score:weights mm_std -score::hbond_His_Phil_fix -no_his_his_pairE -nstruct 1 -out:pdb_gz -overwrite

Example command line for refinement of dependent designs:

fixbb.linuxgccrelease -database __ROSETTA_DATABASE__ -s 2YXNA_min.pdb -minimize_sidechains -score::hbond_His_Phil_fix -no_his_his_pairE -ex1 -ex2 -nstruct 100 -resfile __PREFIX__.resfile -overwrite

Culture and selection conditions. Growth media consisted of LB^L (10 g l^{-1} bacto tryptone, 5 g l^{-1} sodium chloride, 5 g l^{-1} yeast extract). Permissive growth media for bipA-dependent auxotrophs was LB^L supplemented with sodium dodecyl sulfate (SDS), chloramphenicol, bipA and arabinose. Non-permissive media lacked bipA and arabinose. The following selective agents, nutrients and inducers were used when indicated: chloramphenicol (20 μg ml^{-1}), kanamycin (30 μg ml^{-1}), spectinomycin (95 μg ml^{-1}), tetracycline (12 μg ml^{-1}), zeocin (10 μg ml^{-1}), gentamycin (5 μg ml^{-1}), SDS (0.005% w/v), vancomycin (64 μg ml^{-1}), colicin E1 (ColE1; ~10 μg ml^{-1}), DAP (75 μg ml^{-1}), thymidine (100 μg ml^{-1}), bipA (10 μM), glucose (0.2% w/v), pyruvate (0.2% w/v), succinate (0.2% w/v), arabinose (0.2% w/v), anhydrotetracycline (30 ng μl^{-1}). For strains adk.d6_tyrS.d8_bipARS.d7 and adk.d6_ tyrS.d8_int permissive media contained 100 μM bipA. Permissive media for metabolic auxotrophs is LB^L supplemented with 75 μg ml^{-1} DAP and 100 μg ml^{-1} thymidine. TolC selections (SDS) and counter selections (colicin E1) were performed as previously described[33]. Tdk selections used LB^L supplemented with 20 μg ml^{-1} 2′-deoxy-5-fluorouridine and 100 μg ml^{-1} deoxythymidine, and counter selections used LB^L supplemented with 5 μM azidothymidine.

Strain engineering. Two strategies were undertaken to engineer redesigned essential proteins. Strains adk.d4, adk.d6, alaS.d5, holB.d1, metG.d3 and pgk.d4 were generated by performing CoS–MAGE[14] with designed single stranded oligonucleotide pools (Supplementary Table 1) and *tolC* co-selection[14]. Recombined populations were plated on permissive media, and then replica plated on non-permissive media to screen for bipA-dependent clones. Top candidates were identified by kinetic growth monitoring (Biotek H1 or H4 plate reader) of 10–40 bipA-dependent clones in permissive and non-permissive liquid growth media. Strains showing robust growth in permissive media and little to no growth in non-permissive media were carried forward. The tyrS.d6, tyrS.d7 and tyrS.d8 gene variants were constructed by PCR amplification of the *E. coli* MG1655 *tyrS* gene with mutagenic primers, followed by full-length Gibson assembly[34] (Supplementary Tables 1 and 3) and λ Red recombineering[35,36]. Strains tyrS.d6, tyrS.d7, tyrS.d8, adk.d6_tyrS.d6, adk.d6_tyrS.d7 and adk.d6_tyrS.d8 were produced by (1) deleting the endogenous *tdk* gene from C321.ΔA, (2) replacing the endogenous *adk* and *tyrS* genes with their codon-shuffled variants (*adk*(recode)-*tdk* and *tyrS*(recode)-*tdk*, Supplementary Table 3) transcriptionally fused to *tdk*, and (3) replacing the fusion cassettes with the adk.d6, tyrS.d6, tyrS.d7, or tyrS.d8 variants. Variants of adk.d6, tyrS.d7 and tyrS.d8 containing leucine and tryptophan at the bipA position were constructed by MAGE with oligonucleotides containing the appropriate mutations and clonal populations were isolated on LB^L plates lacking bipA and arabinose. Triple-enzyme auxotrophs were created by replacing *asd* with a *Δasd::spec^R* cassette. We reactivated mismatch repair using mutS_null_ revert-2* in the *pgk*, *adk* and *tyrS* single-enzyme auxotrophs and all of the multi-enzyme auxotrophs. For construction of the quadruplet tRNA$_{bipA}$ (Supplementary

Discussion) QuikChange was used to replace the CUA anticodon with UCUA. Quadruplet versions of adk.d6 and tyrS.d8 with UAGA at the bipA positions were constructed by PCR and Gibson assembly followed by λ Red-mediated recombination into the genome as described above. All genotypes (Supplementary Table 3) were confirmed using mascPCR[37] and Sanger sequencing using primers from Supplementary Table 1.

Strain doubling time analysis. Strain doubling times were calculated as previously described[9]. Briefly, cultures were grown in flat-bottom 96-well plates (150 μl LB^L, 34 °C, 300 r.p.m.). Kinetic growth (OD$_{600}$) was monitored on a Biotek H1 plate reader at 5-min intervals. Doubling times were calculated by $t_{double} = \Delta t \times \ln(2)/m$, where $\Delta t = 5$ min per time point and m is the maximum slope of $\ln(OD_{600})$ calculated from the linear regression through a sliding window of 5 contiguous time points (20 min intervals). For escapee strains exhibiting growth rates that were too slow for this analysis, doubling times were calculated by $t_{double} = \Delta t \times \ln(2)/\ln(P_2/P_1)$, where Δt represents sliding windows of 15 min and P_2/P_1 represents initial/final OD$_{600}$ values for the window. Strains that exhibited doubling times greater than 900 min and/or maximum OD$_{600}$ values less than 0.2 after the specified culture duration were considered to exhibit no growth ('none observed') for the given conditions. Improved aeration doubling times for strains adk.d6_tyrS.d8_int and adk.d6_tyrS.d8_bipARS.d7 were obtained by growing strains in 3 ml LB^L in 28 ml culture tubes (three tubes dedicated for each time point), measuring OD$_{600}$ of three technical replicates in 1 cm cuvettes in a spectrophotometer (Beckman DU640) at 20 min intervals for 3.66 h, and determining the slope of the log transformed data over each 40 min window. Least doubling time (20 min × ln(2)/slope) and corresponding R^2 values are reported (Supplementary Table 4).

Expression and purification of tyrS.d7. C321.ΔA cells were grown to mid-log phase in LB^L and co-electrotransformed with 5 ng each of plasmid pEVOL-bipA[10] and an additional plasmid containing full-length tyrS.d7 as an amino-terminal GST fusion under an anhydrotetracycline (aTc)-inducible promoter. After 90 min of recovery cells were plated on LB^L agar supplemented with chloramphenicol and kanamycin. Single colonies were used to inoculate 2 ml starter cultures of LB^L supplemented with chloramphenicol and kanamycin that were grown overnight at 34 °C. Saturated overnight growths were diluted 1:100 into six 1 l cultures containing LB^L supplemented with chloramphenicol and kanamycin, which were grown at 34 °C with shaking at 250 r.p.m. to an OD$_{600}$ of 0.6. The temperature was then reduced to 18 °C, and bipA was added to a final concentration of 500 μM. After an additional 60–90 min aTc and arabinose were added to final concentrations of 30 ng ml^{-1} and 0.2%, respectively. After 24 h of expression, cells were harvested by centrifugation at 10,000g and snap frozen in a dry ice and ethanol bath. Approximately 10 g of thawed cell pellet was suspended in 100 ml of Buffer A (20 mM Tris–HCl (pH 7.2), 500 mM NaCl, and 5% (v/v) glycerol) supplemented with 1 mg ml^{-1} lysozyme. After sonication (6 cycles of 30 s each), the cell lysate was centrifuged at 20,000g for 20 min at 4 °C, and the supernatant was mixed with 5 ml of polyethyleneimine (pH 7.9) on ice. After centrifugation again at 20,000g for 10 min at 4 °C, the supernatant was filtered through a 0.45 μm PVDF membrane, and suspended with 3 ml of glutathione sepharose 4B beads (GE Healthcare Life Sciences). The beads were extensively washed with Buffer A supplemented with 1 mM dithiothreitol (DTT) and incubated with 120 units of PreScission protease (GE Healthcare Life Sciences) at 5 °C for 16 h. The untagged protein was eluted and dialysed against Buffer B (20 mM Tris–HCl (pH 7.5), 50 mM NaCl, 10 mM 2-mercaptoethanol and 5% (v/v) glycerol) at 4 °C. The redesigned enzyme was concentrated to approximately 5.5 mg ml^{-1}.

Determination of the tyrS.d7 crystallographic structure. One-microlitre drops of protein were mixed with an equal volume of a reservoir solution containing 0.1 M sodium malonate (pH 5.5) and 18% (w/v) polyethylene glycol 3350. Crystals were grown at room temperature via hanging drop vapour phase diffusion, and then were transferred into 0.1 M sodium malonate (pH 5.5), 25% (w/v) polyethylene glycol 3350 and 15% ethylene glycol. Crystals were frozen in liquid nitrogen, and diffraction data were collected at the Advanced Light Source (ALS) beamline 5.0.1. The data were processed using the HKL2000 package[38]. The crystal structure of a truncated E. coli TyrS variant protein (PDB code 2YXN) was used as a search model in molecular replacement. The crystallographic model was built using COOT[39], refined using REFMAC5 and Crystallography and NMR system (CNS)[40], and deposited in the RCSB Protein Data Bank (PDB code 4OUD). Statistics of the data collection and refinement are provided in Extended Data Table 1.

Mass spectrometry of NSAA-dependent enzymes. Strains adk.d6, tyrS.d7 and tyrS.d8 were grown to mid-log phase in 10 ml of permissive media. Cell pellets were obtained and soluble lysate fractions were collected as above. Samples were normalized to 250 μg (adk.d6) or 50 μg (tyrS.d7 and tyrS.d8) total protein content and resolved by SDS–PAGE. Gel slices from each strain containing the enzyme (resolved by size comparison to a known standard) were digested with trypsin. Peptide sequence analysis of each digestion mixture was performed by microcapillary reversed-phase high-performance liquid chromatography coupled with

nanoelectrospray tandem mass spectrometry (μLC-MS/MS) on a LTQ-Orbitrap Elite mass spectrometer (ThermoFisher Scientific, San Jose, CA). The Orbitrap repetitively surveyed an m/z range from 395 to 1,600, while data-dependent MS/MS spectra on the 20 most abundant ions in each survey scan were acquired in the linear ion trap. MS/MS spectra were acquired with relative collision energy of 30%, 2.5-Da isolation width, and recurring ions dynamically excluded for 60 s. Preliminary sequencing of peptides was facilitated with the SEQUEST algorithm with a 30 p.p.m. mass tolerance against the Uniprot Knowledgebase E. coli K-12 reference proteome supplemented with a database of common laboratory contaminants, concatenated to a reverse decoy database. Using a custom version of Proteomics Browser Suite (PBS v.2.7, ThermoFisher Scientific), peptide-spectrum matches were accepted with mass error <2.5 p.p.m. and score thresholds to attain an estimated false discovery rate of ~1%.

Western blot analysis of tyrS.d7 variant GST fusions. Cell pellets for all variants were obtained as described above, with an expression culture volume of 10 ml. Cells were lysed using B-PER Bacterial Protein Extraction Reagent, lysozyme (100 mg ml^{-1}), DNaseI (5,000 U ml^{-1}), and Halt Protease Inhibitor Cocktail (all Thermo Scientific) according to the manufacturer's specifications. Lysates were centrifuged at 15,000g for 5 min and the soluble fractions were collected. Protein concentration was determined fluorometrically using the Qubit Protein Assay Kit (Life Technologies). Lysates were normalized to 5-μg samples, resolved by SDS–PAGE, and electro-blotted onto PVDF membranes (Life Technologies, number IB24002). Western blotting was performed using an anti-GST mouse monoclonal primary antibody (Genscript, number A00865-40) and anti-GAPDH mouse monoclonal loading control antibody (Thermo Scientific, number MA5-15738) followed by secondary binding to a HRP-conjugated anti-mouse antibody (Thermo Scientific, number 35080). Samples were imaged by luminol chemiluminescence on a ChemiDoc system (BioRad) and protein content was quantified by densitometry and normalized to GAPDH.

Solid media escape assays for natural metabolic and synthetic auxotrophs. All strains were grown in permissive conditions and harvested in late exponential phase. Cells were washed twice in LB^L and resuspended in LB^L. Viable c.f.u. were calculated from the mean and standard error of the mean (s.e.m.) of three technical replicates of tenfold serial dilutions on permissive media. Three technical replicates were plated on non-permissive media and monitored for 7 days. The order of magnitude of cells plated ranged from 10^2 to 10^9 depending on the escape frequency of the strain. Synthetic auxotrophs were plated on two different non-permissive media conditions: SC, LB^L with SDS and chloramphenicol (Supplementary Table 4); and SCA, LB^L with SDS, chloramphenicol and 0.2% arabinose (Supplementary Table 5). Metabolic auxotrophs were plated on LB^L for non-permissive conditions (Supplementary Table 11). If synthetic auxotrophs exhibited escape frequencies above the detection limit (lawns) on SC at day 1, 2 or 7 (alaS.d5, metG.d3, tyrS.d7), escape frequencies for those days were calculated from additional platings at lower density. Additional platings at higher density were also used to obtain day 1 and day 2 escape frequencies for pgk.d4 on SC. The s.e.m. $S_{\bar{X}}$ across technical replicates of the cumulative escape frequency v scored for a given day was calculated as: $S_{\bar{X}} = v \sqrt{\left(\frac{S_{\bar{X}}\tau}{\tau}\right)^2 + \left(\frac{S_{\bar{X}}n}{n}\right)^2}$, where τ is the mean number of c.f.u. plated, $S_{\bar{X}}\tau$ is the s.e.m. of c.f.u. plated, n is the mean cumulative colony count up to the given day, and $S_{\bar{X}}n$ is the s.e.m. of the cumulative colony count up to the given day. If synthetic auxotroph escapees emerged on SC, three clones were isolated, their growth rates were calculated as described above, and the doubling time of the fastest escapee was recorded (Supplementary Table 4).

Site saturation mutagenesis at designed UAG positions. To site-specifically replace UAG with all other codons, we used MAGE oligonucleotide pools that exactly matched the sequence of the bipA-dependent gene except that the UAG was replaced by all 64 NNN codons (Supplementary Table 1). This allowed us to assess which canonical amino acid substitutions resulted in the best survival of synthetic auxotroph escapees. Although some of these amino acid substitutions may be unlikely to be evolutionarily sampled (evolution will favour amino acids with many tRNA gene copies and whose cognate codons are a single mutation from UAG[41]), this unbiased strategy avoided missing mechanisms of tRNA suppression that are not yet characterized. Immediately after introducing NNN codon diversity via MAGE[16], we recovered the cell populations in 1 ml of LB^L without supplementing antibiotics, arabinose, or bipA. At this point, functional proteins using bipA for proper expression, folding and function are still present in the cell, but protein turnover eventually replaces the bipA-dependent proteins with bipA-independent variants in which the UAG codon is replaced by one of the 64 codons. This in turn provides a strong selection for canonical amino acids that can replace bipA and maintain protein function. Samples of the population were taken at five time points after electrotransformation to track the population dynamics—after 1 h, 100 μl of culture was centrifuged at 16,000g, resuspended in 20 μl distilled water (dH$_2$O), and frozen

at –20 °C (time point 1); 2 ml of LB^L was added to the culture and then growth was allowed to proceed for 3 more hours before 100 μl of culture was centrifuged at 16,000g, resuspended in 20 μl dH$_2$O, and frozen at –20 °C (time point 2); the remaining culture was grown overnight to confluence after which 500 μl of culture was centrifuged at 16,000g, resuspended in 500 μl dH$_2$O, and frozen at –20 °C (time point 3); 30 μl of confluent culture was diluted into 3 ml of fresh LB^L and re-grown to confluence after which 500 μl of culture was centrifuged at 16,000g, resuspended in 500 μl dH$_2$O, and frozen at –20 °C (time point 4); finally, 30 μl of confluent culture was diluted into 3 ml of fresh LB^L and re-grown to confluence after which 500 μl of culture was centrifuged at 16,000g, resuspended in 500 μl dH$_2$O, and frozen at –20 °C (time point 5). After sampling was complete, we had obtained five time points from eight strains amounting to 40 total samples. Population dynamics were analysed by next-generation sequencing.

Next-generation sequencing of populations with degeneracy introduced at UAG positions. We designed custom primers to amplify ~127–146 base pairs (bp) surrounding the UAG codon of each variant and to add Illumina adapters and barcodes for sequencing. In order to reduce primer dimers, we redesigned the P5 primer binding sequence (Sol-P5_alt-PCR, Supplementary Table 1). We used PCR to introduce Illumina sequencing primer binding sites separated from the target amplicon by a 4–6 bp 'heterogeneity spacer' that allows low diversity Illumina libraries to be sequenced out of phase[42] (Supplementary Table 1). We estimate that ~10^6 cells (1 μl of a confluent culture containing ~10^9 cells per ml) were assayed at each time point. This PCR was performed in 20 μl reactions containing 10 μl of KAPA HiFi HotStart ReadyMix, 9 μl of dH$_2$O, 0.5 μl of each 20 μM primer, and 1 μl of template cells. Thermocycling (BioRad C1000 thermocycler) involved heat activation at 95 °C for 3 min, followed by 30 cycles of denaturation at 98 °C for 20 s, annealing at 62 °C for 15 s, and elongation at 72 °C for 30 s with a final elongation for 1 min (PCR1). PCR1 products (20 μl) were purified with magNA beads (40 μl)[43] and eluted in 20 μl of dH$_2$O. A second PCR (PCR2) amplification introduced Illumina adapters tagged with a unique 6 bp barcode (on the P7 adaptor) for each sample and time point. The PCR2 thermocycling and purification protocols were identical to those of PCR1 except that the products from PCR1 were used as template and different primers were used. The final DNA libraries were checked on a 1.5% w/v agarose gel and quantitated using a NanoDrop ND-1000 spectrophotometer. Equimolar samples of all 40 libraries were combined in a single tube and sequenced using a SE50 kit on an Illumina MiSeq (Dana Farber Cancer Institute Molecular Biology Core Facility). The P7 and Index1 reads were performed with standard sequencing primers, whereas the P5 read was sequenced with a custom primer (Sol-P5_alt-PCR, Supplementary Table 1).

Sequencing analysis of populations with NNN degeneracy at UAG positions. A simple Python script was written to tally each of the 64 UAG→NNN codon mutations and 21 amino acid/translational stop substitutions. We discarded all reads that were too short to discern the NNN codon. For all other reads, a constant seed sequence was indexed within the read, and the NNN codon was located based on proximity to this known seed sequence. The NNN codon identity and translated amino acid identities were stored in dictionaries entitled 'aas' and 'codons', respectively. The dictionaries and code are available together at GitHub (https://github.com/churchlab/NNN_sequencing_scripts).

Shannon entropy calculations. Shannon entropy was calculated using the standard relation $H(X) = -\sum_i P(x_i) \log P(x_i)$.

Whole-genome sequencing analysis of mutagenic escapees. We performed whole-genome sequencing on 20 escapees and their bipA-dependent parental strains. Sequencing libraries were prepared according to ref. 43 and sequenced with 150-bp paired-end reads on an Illumina MiSeq. We used Millstone (http://churchlab.github.io/millstone/) to automatically call single-nucleotide variants from raw fastq data with respect to our starting GRO C321.ΔA (NCBI GenBank Accession CP006698.1). Thus all variant positions are reported relative to the frame of this genome. All variant calls are available on Github (https://github.com/churchlab/dependence/tree/master/supplementary_materials). We then filtered these with custom scripts (https://github.com/churchlab/dependence) to identify alleles involved in hypothetical escape mechanisms: mutations in tRNAs that could lead to UAG suppression, mutations in translation machinery that could increase misincorporation of canonical amino acids, mutations in functionally related genes that could functionally complement the essential gene, and mutations in chaperones or proteases that could stabilize poorly folded Adk and TyrS proteins. Additionally, for strains with adequate coverage we performed *de novo* assembly of unmapped reads to uncover structural variants not reported by Millstone. We used Velvet[44] with a hash length of 21 and the following parameters for the graphing step: -cov_cutoff 20 -ins_length 200 -ins_length_sd 90. We then systematically queried NCBI BLAST to identify each *de novo* sequence and biased the BLAST results to prefer hits against the canonical MG1655 genome so that we could later group contigs by position (https://github.com/churchlab/dependence/blob/master/supplementary

_materials/velvet_contigs_and_BLAST_data_non_permissive_8_strains.csv). Putative Lon insertions were visually confirmed using Millstone's JBrowse portal.

Integration of *bipARS* and tRNA$_{bipA}$. Genomic integration of *bipARS* and tRNA$_{bipA}$ was achieved in two steps by first replacing the endogenous *tdk* gene with the P$_{araBAD}$-inducible *bipARS* gene from pEVOL. Subsequently, the pEVOL tRNA and chloramphenicol resistance gene were inserted immediately downstream of P$_{araBAD}$-*bipARS*. Kapa HiFi Ready Mix was used to amplify each PCR product (see Supplementary Table 1 for primer sequences), and λ Red-mediated recombination was used to introduce the PCR products into the genome. Proper insertion of the desired cassettes was confirmed by PCR using tdk.seq-f and tdk.seq-r. We observed that 10 μM bipA was not adequate to support growth of adk.d6 or tyrS.d8 when *bipARS* was integrated into the genome; however, 100 μM bipA accommodated robust growth of adk.d6_int, tyrS.d8_int and adk.d6_tyrS.d8_int.

Design of the bipARS.d7 bipA-dependent synthetase. We applied our computational second-site suppressor strategy to position V290 of the bipyridylalanyl-tRNA synthetase X-ray structure (PDB code 2PXH, chain A). This position corresponds to bipA303 in our X-ray structure of tyrS.d7 when the structures are superimposed (alignable core backbone root mean square deviation of 3.6 Å). We hypothesized that this position may be amenable to redesign in homologous structures. Six designs covering sequence variability observed in the computational models (Supplementary Table 2) were produced by PCR amplification of *bipARS* with mutagenic primers (Supplementary Table 1) and isothermal assembly into the pEVOL vector, maintaining only the arabinose-inducible copy of *bipARS*. We also included the D286R mutation previously shown to increase synthetase activity[45] in all constructs. Since the *bipARS* designs should require bipA to translate, fold and function, all derived strains were initially co-transformed with a nonreplicating plasmid (pJTE2, R6γ origin of replication) containing a wild-type copy of *bipARS* to jumpstart production of tRNA$_{bipA}$. Designs were co-transformed with the jumpstart plasmid into C321.ΔA, and transformants were then transformed with a previously described GFP reporter plasmid containing a single UAG codon[9] to measure synthetase activity by GFP fluorescence. One design (bipARS.d7; T/A/G/G/A/bipA) produced >5-fold bipA-dependent induction of fluorescence in permissive media but failed to induce any bipA-dependent fluorescence after passaging overnight 1:150 in non-permissive media followed by an identical passage in permissive media. Since any functional synthetase remaining after non-permissive passaging should facilitate exponential production of additional synthetase, this behaviour suggests strong dependence of bipARS.d7 on bipA for translation and folding resulting in total clearance of bipARS.d7 and tRNA$_{bipA}$ after overnight growth in non-permissive conditions. The bipARS.d7/tRNA$_{bipA}$ and jumpstart vectors were co-transformed into C321.ΔA and then adk.d6 and tyrS.d8 were introduced as described above.

Growth competition assays. The assayed single- and double-enzyme synthetic auxotroph escapee strains (pgk.d4 esc. 1, 2 and 3; adk.d6_tyrS.d8 esc. 1, 2 and 3) were transformed with a pZE21 vector[46] bearing mCFP under aTc-inducible control in the multiple cloning site. The parental prototrophic C321.ΔA strain was similarly transformed with an identical vector except that the fluorophore is YFP. Strains were grown to late-exponential phase in LB^L supplemented with antibiotics (SDS, chloramphenicol, kanamycin), inducers (0.1% L-arabinose, 100 ng ml^{-1} aTc) and bipA. Cells were washed twice in M9 salts and adjusted to a cell concentration of roughly 1×10^9 cells per ml. Biological replicates of synthetic auxotroph escapees were mixed with the C321.ΔA strain at a ratio of 100:1 and diluted to a seeding concentration of roughly 2.5×10^7 cells per ml in non-permissive media (LB^L supplemented with SDS, chloramphenicol, kanamycin and aTc). Growth kinetics of the competition mixture were assayed in 200 μl sample volumes on microtitre plates incubated in a Biotek Synergy microplate reader at 34 °C. Cell mixtures were fixed in PBS with 1% paraformaldehyde at time 0 and at 8 h. Fixed cells were run on a BD LSRFortessa cell analyser and populations were binned based on YFP expression level. CFP was not used for species discrimination but rather to maintain consistent fitness costs associated with episomal DNA maintenance and fluorophore expression.

Bacterial lysate growth assays. All strains were grown up in permissive conditions and harvested in late-exponential phase. Cells were washed twice in M9 salts (6 g l^{-1} Na$_2$HPO$_4$, 3 g l^{-1} KH$_2$PO$_4$, 1 g l^{-1} NH$_4$Cl, 0.5 g l^{-1} NaCl) by centrifugation at 17,900g and then diluted 100-fold into LB^L supplemented with 166.66 ml l^{-1} trypsin-digested *E. coli* extract (Teknova catalogue number 3T3900). Growth kinetics were assayed in 200 μl sample volumes on microtitre plates as described above. Three biological replicates were performed by splitting a single well-mixed initial seeding population.

Conjugal escape assays. The conjugal donor population was produced using the Epicentre EZ-Tn5 Custom Transposome kit to insert a mosaic-end-flanked kan^R-oriT cassette into random positions of the *E. coli* MG1655 genome. The population of integrants was plated on LB^L agar plates supplemented with kanamycin. Approximately 450 clones were lifted from the plate and pooled, which corresponds

to one kan^R-oriT per ~10-kilobase pair region of the genome, assuming an equal distribution of transposition across the 4.6-megabase *E. coli* MG1655 genome. The pRK24 conjugal plasmid was conjugated[37] from *E. coli* strain 1100-2 (ref. 47) into the kan^R-oriT donor population. The kan^R-oriT insertion sites were confirmed to be well distributed. In brief, the donor population was sheared on a Covaris E210, end repaired, and ligated to Illumina adapters as described by ref. 43. Genomic sequences flanking the insertion site were amplified using the Sol-P5-PCR primer and a series of nested primers (Supplementary Table 1) that hybridize within the kan^R gene. PCR products corresponding to ~1 kilobase pair were gel purified from the smear and TOPO cloned (Invitrogen pCR-Blunt II-TOPO). Flanking genomic sequences were then identified by Sanger sequencing 96 TOPO clones. Conjugal escape assays were performed as described previously[37] with 50-min and 12-h conjugal duration and a donor:auxotroph ratio of 1:100. Three technical replicates of two biological replicates were performed for all conjugation assays with the exception of the double-enzyme synthetic auxotroph experiments, which were performed with three biological replicates (3 technical replicates each) to produce enough escapees for mascPCR screening. To determine the proportion of the genome overwritten by donor DNA the following numbers of colonies were scored for the 50-min/12-h time points: adk.d6 $n = 51/6$; tyrS.d8 $n = 44/7$; adk.d6_tyrS.d8 $n = 8/59$; adk.d6_tyrS.d8_asd:specR $n = 5/38$. This set omits a small collection of clones that could not be scored due to polyclonality.

Statistics. No statistical methods were used to predetermine sample size.

27. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **33,** D34–D38 (2005).
28. UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41,** D43–D47 (2013).
29. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26,** 689–691 (2010).
30. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319–333 (1998).
31. Zhu, H., Fraczkiewicz, R. & Braun, W. *Solvent Accessible Surface Areas, Atomic Solvation Energies, and Their Gradients for Macromolecules* http://curie.utmb.edu/area_man.html (2012).
32. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97,** 10383–10388 (2000).
33. Gregg, C. J. *et al.* Rational optimization of *tolC* as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* **42,** 4779–4790 (2014).
34. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
35. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
36. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97,** 5978–5983 (2000).
37. Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333,** 348–353 (2011).
38. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (ed Carter, C. W. Jr) 307–326 (Academic, 1997).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
40. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
41. Eggertsson, G. & Soll, D. Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiol. Rev.* **52,** 354–374 (1988).
42. Fadrosh, D. W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2,** 6 (2014).
43. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–946 (2012).
44. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
45. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli. J. Mol. Biol.* **395,** 361–374 (2010).
46. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).
47. Tolonen, A. C., Chilaka, A. C. & Church, G. M. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* **74,** 1300–1313 (2009).
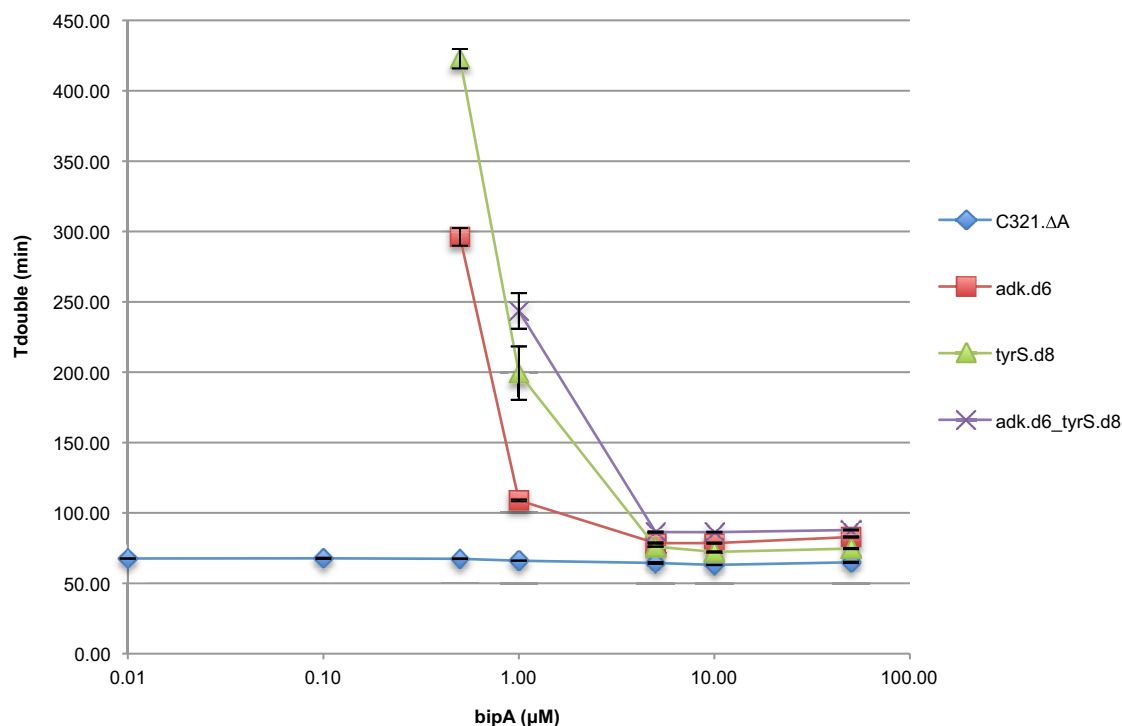
Extended Data Figure 1 | bipA dependence in synthetic auxotrophs.
Prototrophic and synthetic auxotrophic strains were grown in titrations of bipA and monitored in a microplate reader (Methods). Media for all bipA concentrations contained SDS, chloramphenicol and arabinose. Doubling times for three technical replicates are shown. Positive and negative error bars are s.e.m. Growth was undetectable for synthetic auxotrophs at 0.00 μM, 0.01 μM and 0.10 μM bipA, as well as 0.50 μM bipA for adk.d6_tyrS.d8.
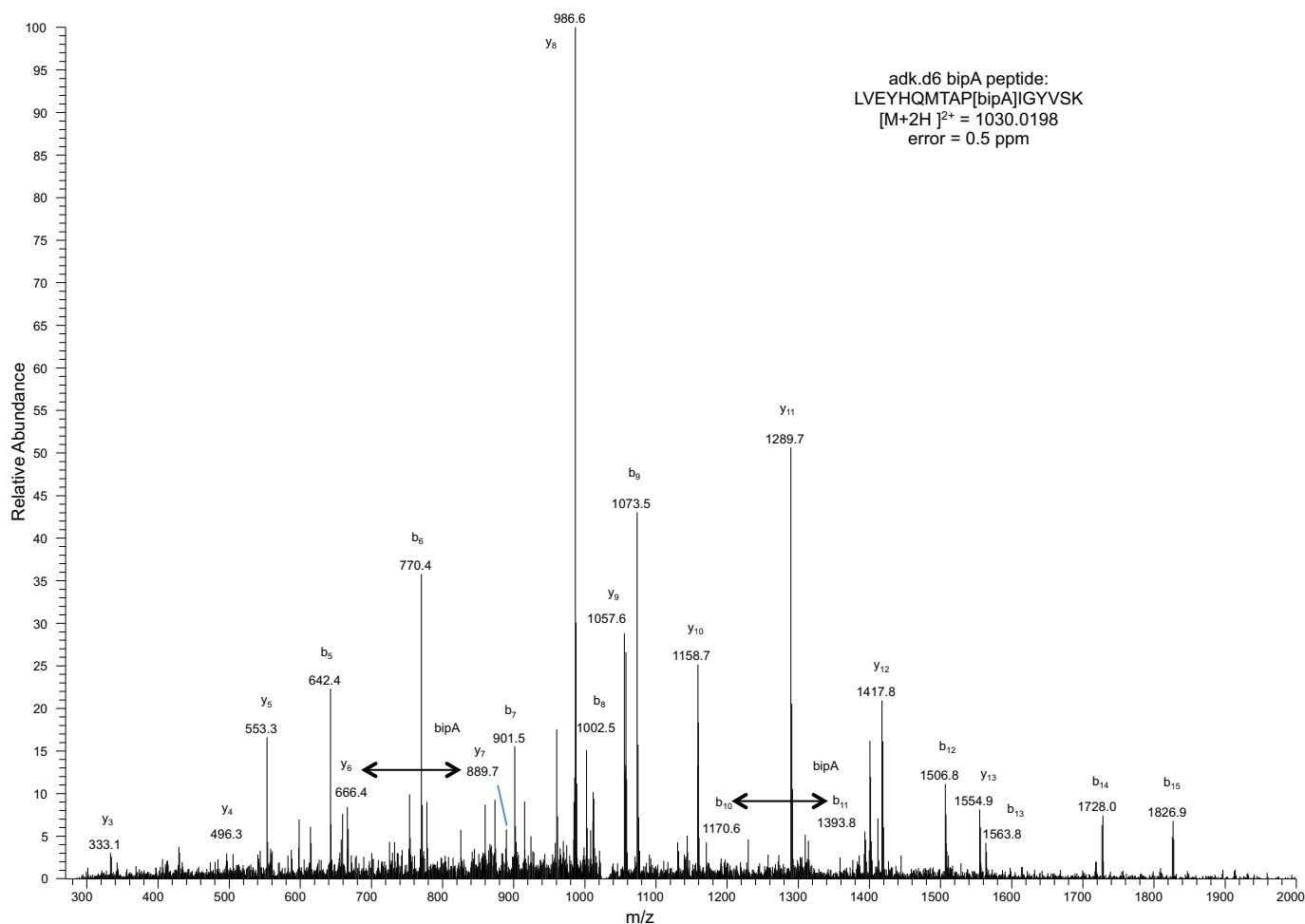
Extended Data Figure 2 | Mass spectrometry of NSAA-dependent enzymes.
Mass spectrometry was performed and peptide spectrum matches (PSMs) were obtained as described in the Methods. Data sets were culled of minor contaminant PSMs and re-searched with SEQUEST against adk.d6, tyrS.d7 and tyrS.d8 sequences without taking into account enzyme specificity. To interrogate the sequences for bipA, tryptophan and leucine, the amino acid at the bipA position was given the mass of leucine and searches were performed with differential modifications of +110.01565 and +72.99525 to account for the masses of bipA and tryptophan, respectively. In all samples, only bipA, and not leucine or tryptophan, was detected at these positions. The PSM for adk.d6 is shown. Peptides observed to contain bipA are LVEYHQMTAP[bipA]IGYVSK (adk.d6), AQYV[bipA]AEQVTR (tyrS.d7) and AQYV[bipA]AEQATR (tyrS.d8).
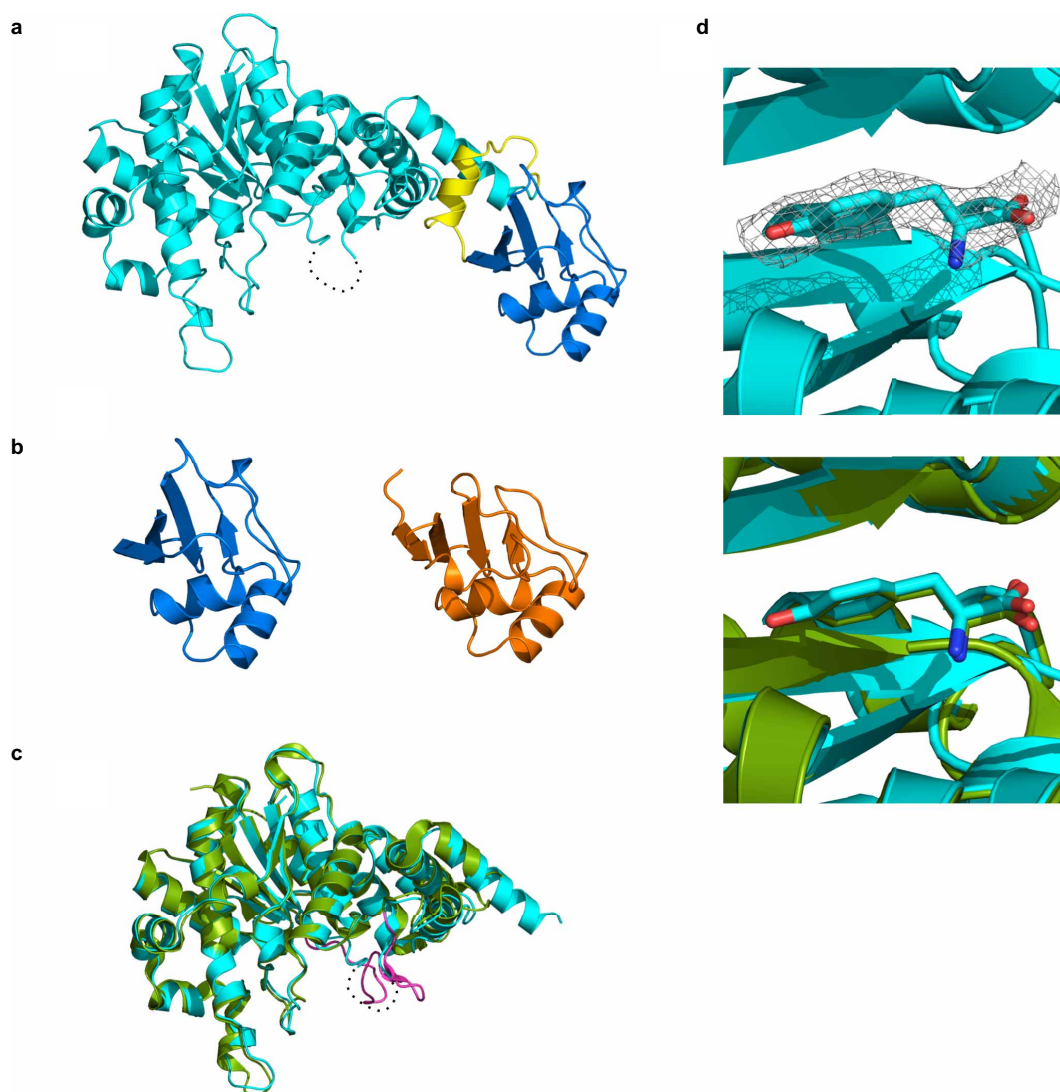
Extended Data Figure 3 | Crystal structure of tyrS.d7. a, Overall structure of the redesigned enzyme. The N-terminal domain (residues 4–330) that catalyses tyrosine activation, the carboxy-terminal tRNA-binding domain (residues 350–424) and their connecting region are coloured cyan, blue and yellow, respectively. The residues 232–241 are disordered (dash line). **b,** Comparison between the C-terminal tRNA recognition domains of tyrS.d7 (blue) and of *Thermus thermophilus* TyrS (orange; PDB code 1H3E). The residues 352–442 of the hyperthermophilic TyrS are shown. **c,** The N-terminal domain of the engineered protein is superposed on the crystal structure of its parental enzyme (green; PDB code 1X8X). The KMSKS loop of the parental enzyme is highlighted in magenta. **d,** Tyrosine molecule bound to tyrS.d7. An electron density map of L-tyrosine is shown as a grey mesh ($2F_o - F_c$ contoured at 1.2σ; top panel). A tyrosine and the surrounding protein fold of tyrS.d7 (cyan) are very similar to those of the wild-type TyrS structure (green; bottom panel).
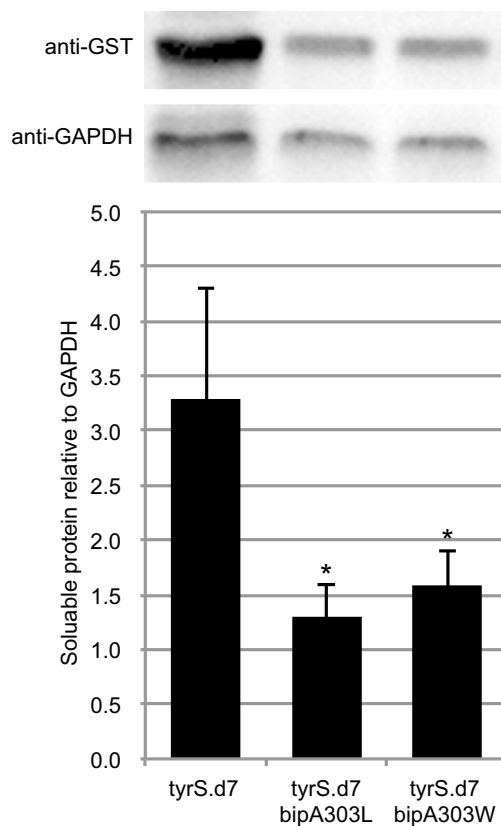
Extended Data Figure 4 | **Western blot analysis of tyrS.d7 variants.** Variants of tyrS.d7 with leucine or tryptophan at the bipA position were expressed as GST fusions under identical conditions and analysed by western blot (Methods). Soluble protein content was quantified by densitometry and normalized to GAPDH. Mutating bipA to leucine or tryptophan reduced soluble TyrS levels by 2.5- or 2.1-fold, respectively (*$P < 0.05$ by two-tailed unpaired Student's t-test with unequal variances). Three technical replicates were performed; a representative image is shown. Positive error bars are s.e.m.

Conf. = Confluent culture
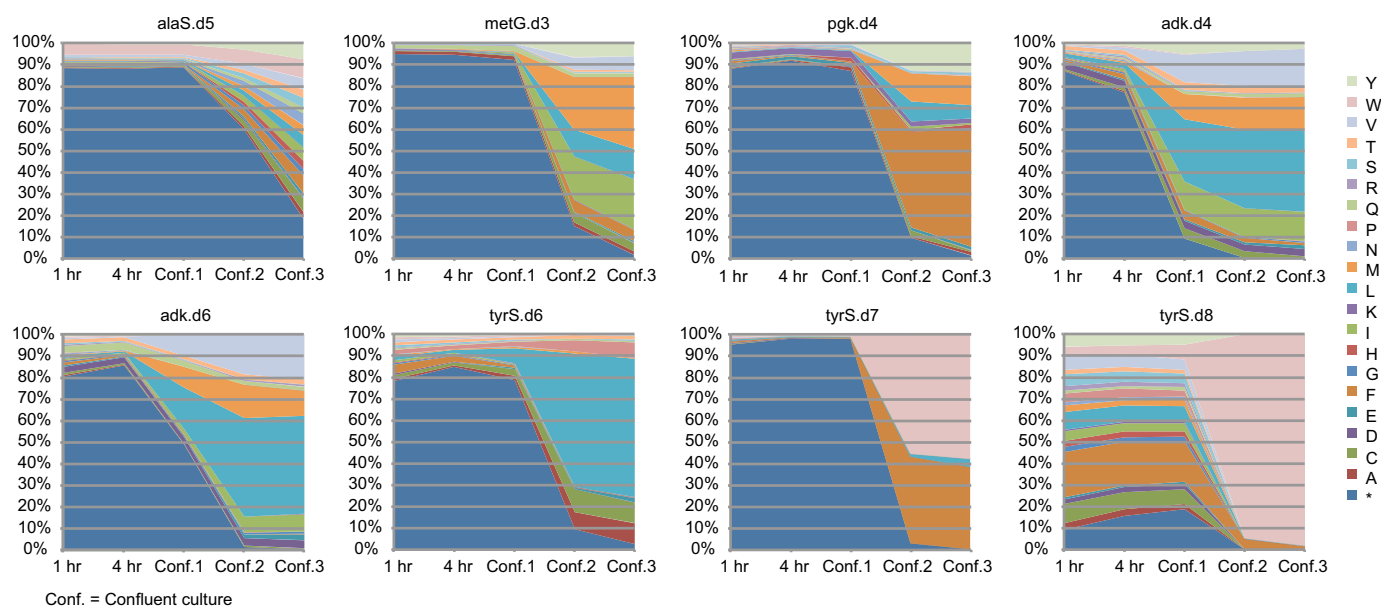
Extended Data Figure 5 | Population selection dynamics for canonical amino acid substitutions at designed UAG positions. For each plot, degenerate MAGE oligonucleotides were used to create a population of cells in which the UAG codon was mutated to all 64 codons. Codon substitutions leading to survival in the absence of bipA were selected by growth in LB^L media without bipA and arabinose supplementation. Aliquots of the culture population were taken at 1 h, 4 h, confluence 1 (once the culture reached confluence), confluence 2 (after regrowth of a 100-fold dilution of confluence 1) and confluence 3 (after regrowth of a 100-fold dilution of confluence 2). The amino acid identity at the bipA position was probed by targeted Illumina sequencing. Residual bipA-containing proteins were expected to remain active until intracellular protein turnover cleared them from the cell, making the 1-h time point a reasonable representation of initial diversity present in the population. These data show the relative fitness of amino acid substitutions in a given protein variant; relative fitness across multiple protein variants cannot be accurately assessed from these data.
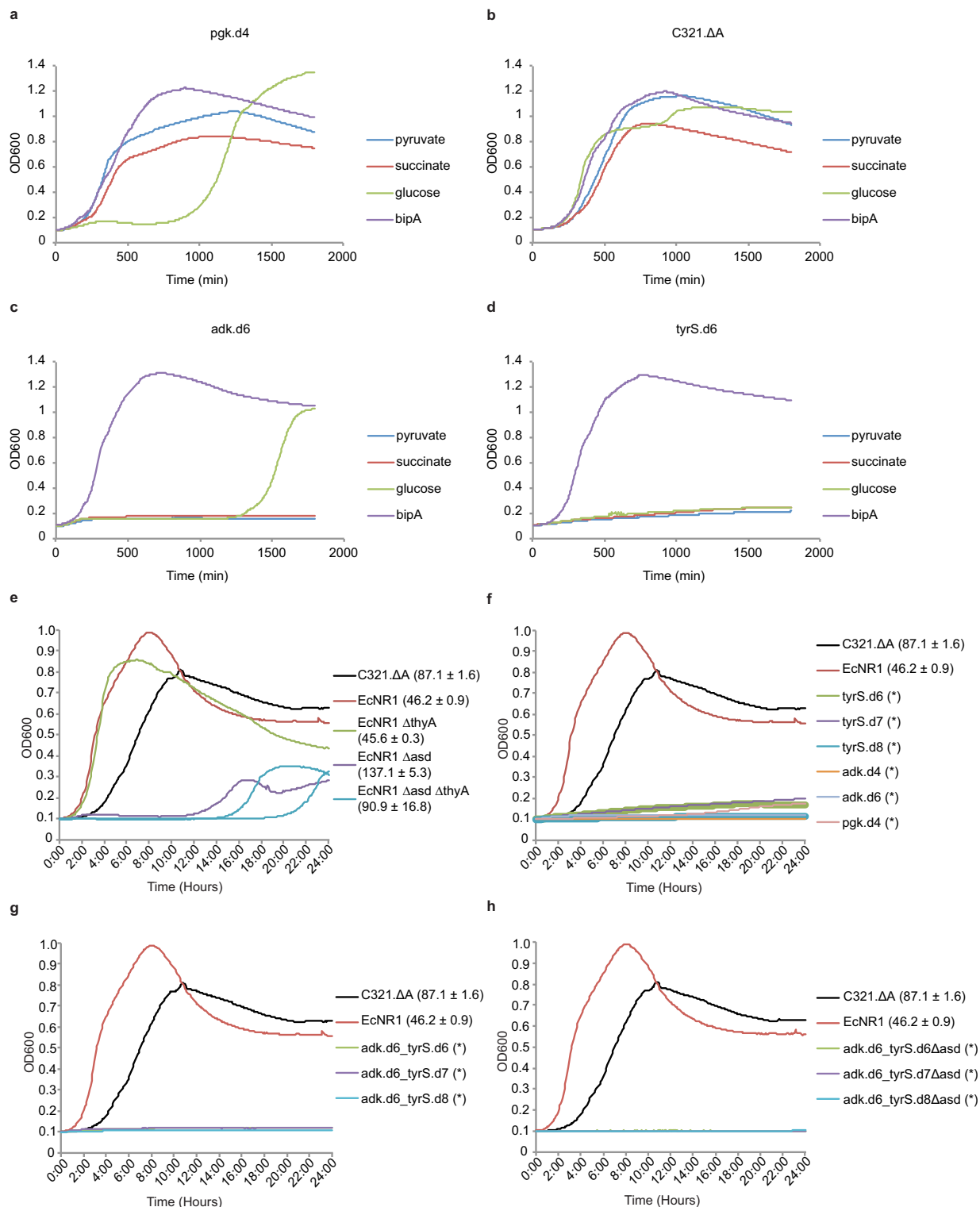
Extended Data Figure 6 | Natural metabolites can circumvent auxotrophies. a–d, Synthetic auxotrophs of *pgk* can be complemented by pyruvate or succinate. Strains were cultured in LB^L in the presence of pyruvate, succinate, glucose or bipA (10 µM) and monitored by kinetic growth. The single-enzyme synthetic auxotroph pgk.d4 (**a**) grows similarly to prototrophic C321.ΔA (**b**) in the presence of pyruvate and succinate, but not glucose. Synthetic auxotrophs of *adk* (**c**) and *tyrS* (**d**) grow robustly in bipA but cannot be complemented by pyruvate or succinate. Growth of pgk.d4 and adk.d6 in glucose after 1,000 min is due to mutational escape (loss of bipA dependence). **e,** The synthetic auxotroph parental strain (C321.ΔA), a second prototrophic MG1655-derived strain (EcNR1), and three natural auxotroph derivatives of EcNR1 were grown in LB^L supplemented with 166.66 ml l^{-1}

bacterial lysate (Teknova). Growth curves are shown with doubling times ± one standard deviation of three technical replicates next to the labels. The conditions fully complement the metabolic auxotrophy of EcNR1.*ΔthyA*, which doubles as robustly as prototrophic EcNR1. Strains lacking the *asd* gene (EcNR1.*Δasd* and the EcNR1.*ΔasdΔthyA* double knockout) show more impairment but enter exponential growth with doubling times of 91 to 137 min, respectively. **f, g,** Single- (**f**) and double-enzyme (**g**) synthetic auxotrophies are not complemented by natural products in rich media or bacterial lysate. **h,** When the *Δasd* auxotrophy is combined with double-enzyme synthetic auxotrophies the natural products are no longer sufficient to support growth. No growth is indicated by an asterisk in **f–h**.

a

b
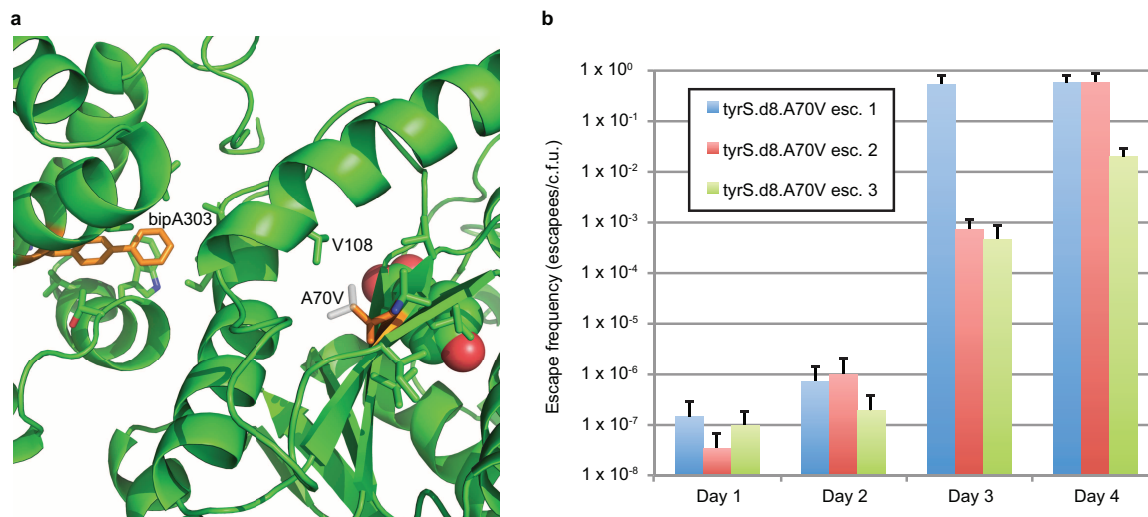
Extended Data Figure 7 | Analysis of the A70V mutation as an escape mechanism for tyrS.d8. a, The X-ray structure of tyrS.d7 is shown; tyrS.d8 varies by the single mutation V307A. BipA303, A70 and their neighbouring side chains are shown in stick representation, with bipA303 and A70 coloured orange. The bound tyrosine substrate is shown in spacefill. The A70V mutation (white sticks) may stabilize the catalytic domain when bipA is replaced by natural amino acids by tightly packing with neighbouring side chains including V108. **b**, Escape frequencies on non-permissive media for three separately constructed tyrS.d8 A70V strains are shown for days 1 through 4. Although escapees are growth-impaired in the absence of bipA (Supplementary Table 10), all cells form colonies after 5 days, suggesting that A70V confers 100% survival on non-permissive media. Positive error bars indicate s.e.m.
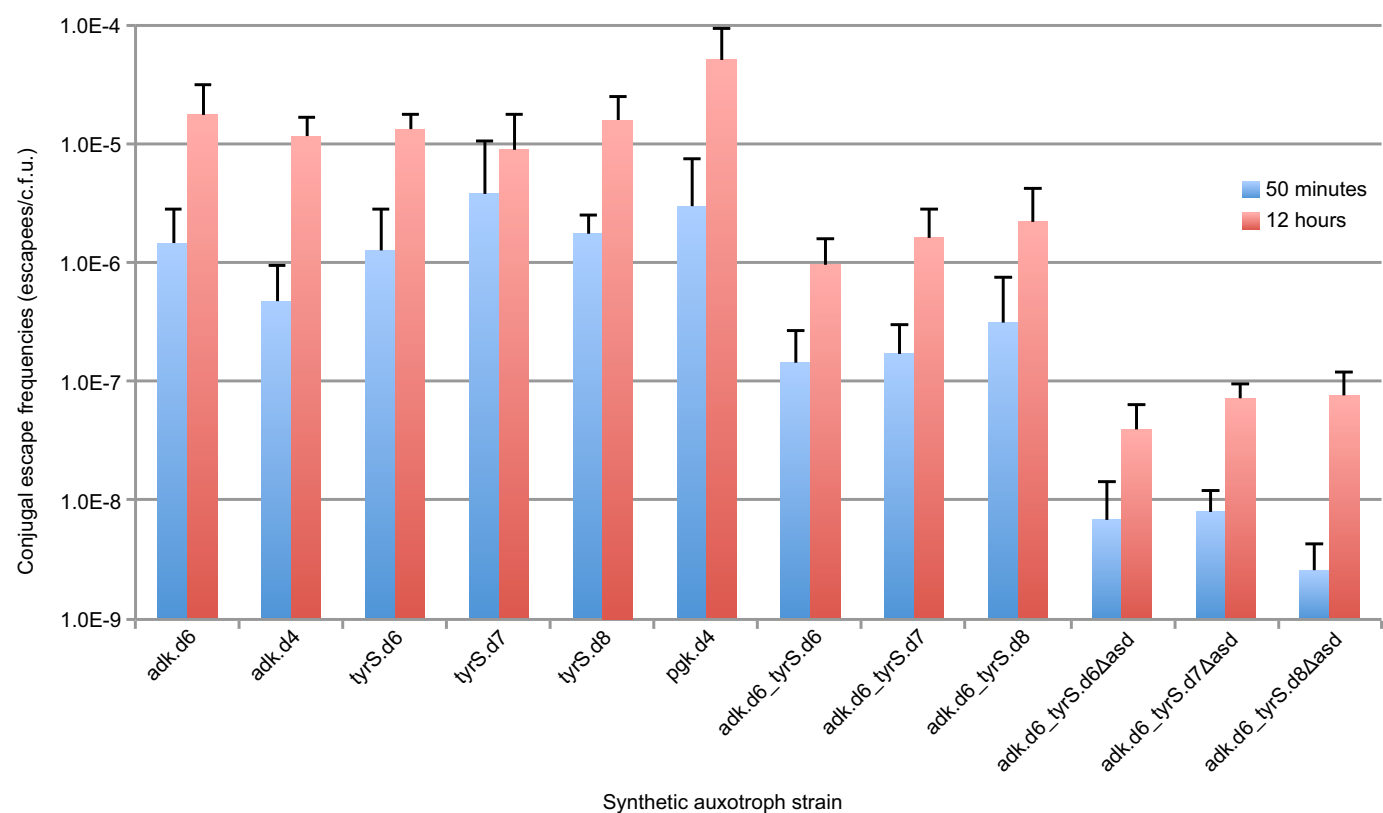
Extended Data Figure 8 | Conjugal escape frequencies of synthetic auxotrophs. Single-, double- and triple-enzyme auxotrophs were assayed to determine the frequency of escape by HGT and recombination from a prototrophic donor as described in the Methods. These results highlight the benefit of having multiple auxotrophies distributed throughout the genome. Notably, scaling from a single synthetic auxotrophy to three distributed auxotrophies results in a reduction of conjugal escape by at least two orders of magnitude. Positive error bars indicate standard deviation.

Extended Data Table 1 | Data collection and refinement statistics

	tyrS.d7
Data collection	
Space group	$P\,12_11$
Cell dimensions	
a, b, c (Å)	81.3, 67.2, 90.7
a, b, g (°)	90.0, 102.6, 90.0
Resolution (Å)	50.0 - 2.65 (2.74-2.65) *
R_{sym} or R_{merge}	0.074 (0.497)
I/sI	29.2 (4.65)
Completeness (%)	99.0 (98.4)
Redundancy	7.6 (7.7)
Refinement	
Resolution (Å)	45 – 2.65
No. reflections	26407
$R_{work}/\,R_{free}$	0.222/0.306
No. atoms	
Protein	6038
Ligand/ion	13
Water	57
B-factors	
Protein	58.66
Ligand/ion	52.10
Water	48.24
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.530

The data were collected using a single crystal.
* Highest-resolution shell is shown in parentheses.

Extended Data Table 2 | Cost per litre of culture for commonly used NSAAs

NSAA	Vendor	Name at vendor	CAS#	MW	Cat# for 1g	Price of 1g	Optimal conc. (mM)	Cost per liter of culture
pAcF	peptech	L-4-Acetylphenylalanine	122555-04-8	207.23	AL624-1	$500.00	1.0	$103.62
pAzF	Bachem	H-4-Azido-Phe-OH	33173-53-4	206.2	F-3075.0001	$285.00	5.0	$293.84
pCNF	peptech	L-4-Cyanophenylalanine	167479-78-9	190.2	AL240-1	$150.00	1.0	$28.53
bpa	peptech	L-4-Benzoylphenylalanine	104504-45-2	269.3	AL660-1	$100.00	1.0	$26.93
napA	peptech	L-2-Naphthylalanine	58438-03-2	215.25	AL121-1	$80.00	1.0	$17.22
bipA	peptech	L-4,4'-Biphenylalanine	155760-02-4	241.29	AL506-1	$150.00	0.1	$3.62
pIF	peptech	L-4-Iodophenylalanine	24250-85-9	291.09	AL261-1	$40.00	1.0	$11.64
bipyA	Asis Chem	Bipyridylalanine	custom synthesis	245.282	(25 g price)	$10,000/25g	1.0	$98.11

ARTICLE

Mechanistic insights into the recycling machine of the SNARE complex

Minglei Zhao[1]*, Shenping Wu[2]*, Qiangjun Zhou[1], Sandro Vivona[1], Daniel J. Cipriano[1], Yifan Cheng[2] & Axel T. Brunger[1,3]

Evolutionarily conserved SNARE (soluble N-ethylmaleimide sensitive factor attachment protein receptors) proteins form a complex that drives membrane fusion in eukaryotes. The ATPase NSF (N-ethylmaleimide sensitive factor), together with SNAPs (soluble NSF attachment protein), disassembles the SNARE complex into its protein components, making individual SNAREs available for subsequent rounds of fusion. Here we report structures of ATP- and ADP-bound NSF, and the NSF/SNAP/SNARE (20S) supercomplex determined by single-particle electron cryomicroscopy at near-atomic to sub-nanometre resolution without imposing symmetry. Large, potentially force-generating, conformational differences exist between ATP- and ADP-bound NSF. The 20S supercomplex exhibits broken symmetry, transitioning from six-fold symmetry of the NSF ATPase domains to pseudo four-fold symmetry of the SNARE complex. SNAPs interact with the SNARE complex with an opposite structural twist, suggesting an unwinding mechanism. The interfaces between NSF, SNAPs, and SNAREs exhibit characteristic electrostatic patterns, suggesting how one NSF/SNAP species can act on many different SNARE complexes.

Membrane fusion is essential for many physiological processes in eukaryotic cells, including protein and membrane trafficking, hormone secretion, and neurotransmission[1,2]. The evolutionarily conserved SNARE proteins have a key role in these processes. Specific combinations of SNARE proteins are located on opposite membranes. Upon zippering into a highly stable four-helix bundle—the SNARE complex—they provide the energy for membrane fusion[3,4]. The combinations of SNARE proteins depend on the source of vesicles and the identity of target membranes, but other factors also contribute to the specificity of the membrane targeting[5]. To maintain the pool of individual SNARE proteins, the ATPase NSF, together with SNAPs, disassembles post-fusion and non-productive SNARE complexes into individual protein components using the energy from ATP hydrolysis[6].

NSF was the first protein found to play a key role in eukaryotic trafficking[7,8]. It is a member of AAA+ (ATPases associated with diverse cellular activities) superfamily of ATPases[9], and it forms a homomeric hexamer with a molecular weight of ~500 kDa, with each protomer consisting of an amino-terminal domain (termed N) and two ATPase domains (termed D1 and D2) (Fig. 1a). The D1 domains are responsible for most of the ATPase activity of NSF, whereas the D2 domains are primarily responsible for hexamerization[10]. The N domains are involved in SNAP and possibly SNARE binding[11]. Prior to ATP hydrolysis, the NSF, SNAP, and SNARE complex form the so-called 20S supercomplex[6].

Individual components of 20S supercomplex have been structurally characterized, including the crystal structures of several SNARE complexes[3,12–16], SNAPs[17,18], and the D2 and N domains of NSF[19–22]. Structural studies of full-length NSF and the 20S supercomplex have also been carried out using quick-freeze/deep-etch, negative-staining electron microscopy, and electron cryomicroscopy (cryo-EM)[23–26]. However, due to the low resolution limits of these studies, the detailed molecular architecture of the 20S supercomplex is unknown and critical questions remain to be answered such as: how the adaptor protein SNAP recognizes SNARE complexes, how many SNAPs are involved; how one NSF/SNAP species

disassembles many different SNARE complexes in a promiscuous fashion; as well as the question of what is the molecular mechanism of disassembly.

Here we present the structures of full-length NSF in two different nucleotide states (ATP- and ADP-bound, at 4.2 Å and 7.6 Å resolution, respectively), and structures of two different 20S supercomplexes involving different SNARE substrates determined by single particle cryo-EM, ranging from 7.6 to 8.4 Å resolution. The cryo-EM structures reveal large conformational differences of NSF between ATP- and ADP-bound states, and upon binding to SNAPs and SNAREs. We confirmed by site-directed mutagenesis that the molecular interfaces between SNAPs, SNAREs, and NSF play important roles in disassembly, and propose that recognition at these interfaces is based on characteristic electrostatic patterns. Based on these new insights we speculate about the molecular mechanisms of NSF-mediated SNARE complex disassembly.

ATP- and ADP-bound NSF structures

We developed a new purification scheme to address the heterogeneity of NSF samples caused by mixtures of nucleotide states. In essence, hexameric NSF was monomerized by completely removing the bound nucleotides through size-exclusion chromatography (SEC) (Methods and Extended Data Fig. 1a, b). The resulting NSF protomers could be reassembled into hexamers in the presence of the desired nucleotide. The reassembled ATP- and ADP-bound NSF hexamers were studied by single particle cryo-EM; EDTA was included to prevent hydrolysis. Our reassembled NSF hexamers are functionally active (Extended Data Fig. 1g and Methods).

The reconstruction of ATP-bound NSF is shown in Fig. 1b, c and Extended Data Fig. 2. The reconstruction has an estimated overall resolution[27] of 4.2 Å after masking out flexible N domains (Extended Data Fig. 2e and Extended Data Table 1). All D2 domains, and five out of six D1 domains were well resolved in the final three-dimensional (3D) density map (Fig. 1b and Extended Data Figs 2d and 3a–c). Consistent with the estimated resolution, we observed grooves in α-helices, β-strands

[1]Department of Molecular and Cellular Physiology, Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. [2]Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. [3]Department of Neurology and Neurological Sciences, Department of Structural Biology, Department of Photon Science, Stanford University, Stanford, California 94305, USA.
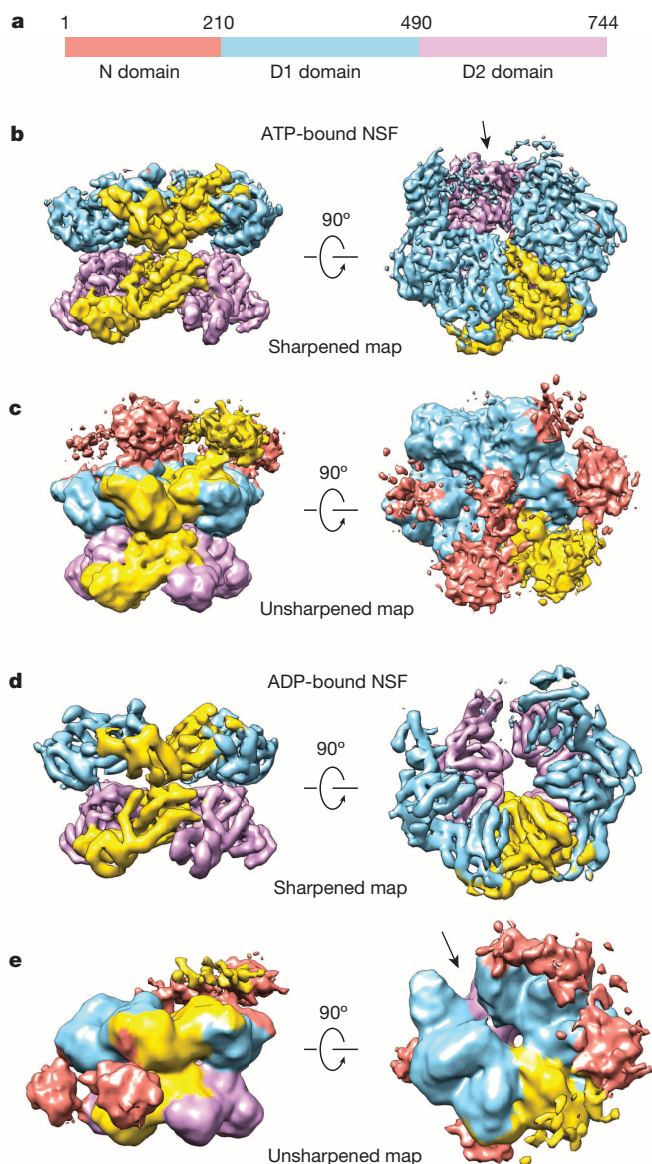*These authors contributed equally to this work.

Figure 1 | 3D density maps of ATP- and ADP-bound NSF. a, Domain diagram of the NSF protomer. **b,** Different views of the sharpened map (6.5σ) of ATP-bound NSF filtered to a resolution of 4.2 Å with each domain colour-coded to match the domain diagram in panel **a**. A single chain of NSF (protomer) is coloured in gold to help with visualization. The density of one D1 domain (subsequently referred to as the D1 domain in chain F) is not well resolved (see black arrow). **c,** Different views of the unsharpened map (1.8σ) of ATP-bound NSF showing the positions of the N domains. **d,** Different views of the sharpened map (6.8σ) of ADP-bound NSF filtered to a resolution of 7.6 Å with each domain colour-coded to match the domain diagram in panel **a**. A single chain of NSF (protomer) is coloured in gold to help with visualization. **e,** Different views of unsharpened map (1.3σ) of ADP-bound NSF showing the positions of the N domains. The gap in the D1 ring is indicated by a black arrow.

within β-sheets, and backbone zigzags corresponding to ~3.8 Å Cα distances, along with densities of some aromatic side chains (Extended Data Fig. 3a, b). The crystal structure of the ATP-bound NSF D2 hexamer[19] was readily docked into the corresponding cryo-EM density, followed by refinement. The densities of the D1 domains were of sufficient quality to build and refine a *de novo* atomic model of the D1 domain with bound ATPs. The D1 domain is a typical AAA+ module with two subdomains (α and α/β) and motifs that are generally found in ATPases (Extended Data Fig. 3d, e). The α2 helix of the D1 domain is bent (Extended Data Fig. 3d), a distinctive feature compared to the straight

α2 helices found in the D2 domain of NSF, as well as in both D1 and D2 domains of the closest related relative, the AAA+ ATPase valosin-containing protein (VCP/p97)[28–30].

The 3D reconstruction of ADP-bound NSF is shown in Fig. 1d, e. The overall estimated resolution[27] is 7.6 Å (Extended Data Fig. 4 and Extended Data Table 1), with well-resolved tubular densities for α-helices (Fig. 1d). To obtain an atomic model of ADP-bound NSF, we docked our cryo-EM structure of the D1 domain protomer (obtained from ATP-bound NSF) and the crystal structure of the D2 domain hexamer[19] into the corresponding cryo-EM densities, followed by refinement. To complete the models, the crystal structure of NSF N domain[21] was docked into the corresponding densities (Methods).

The cryo-EM data sets of the NSF particles used in this study were of sufficient quality to determine and refine 3D reconstructions to high resolution without imposing any symmetry, which turned out to be critical for resolving the NSF N domains and asymmetries in the ATPase rings (Extended Data Fig. 5, see Supplementary Information for a detailed discussion).

Asymmetric features of ATP- and ADP-bound NSF

Both the ATP- and ADP-bound structures of NSF are organized into three layers: two rings consisting of six D2 domains and six D1 domains, respectively, and a layer of six (four) N domains for ATP (ADP)-bound NSF (Figs 1c, e, and 2a, b). For ADP-bound NSF, the remaining two N domains are flipped along the sides of the ATPase rings with well resolved densities compared to the N domains atop the D1 ring, leaving little doubt as regards the identity of these two densities (Fig. 1e and Extended Data Figs 4c, e and 7c).

For ATP-bound NSF, the D2 ring is planar and approximately six-fold symmetric. The D1 ring is reminiscent of a right-handed 'split washer', with each chain stepping up about 5 Å as manifested by the relative positions of the α2 helix in the D1 domains (Fig. 2c and Extended Data Fig. 6a). Chain F (purple) is an exception, which does not step up relative to chain E (blue), but instead slightly steps down towards chain A (red); there is a large step down from chain F to chain A. The D1 domain of chain F was not as well resolved in the density map as the other D1 domains (Fig. 1b), indicating its potential flexibility (Supplementary Video 1). However, the density for this domain is clear enough to indicate that the α subdomain has a different position relative to the α/β subdomain, compared to the other five D1 domains that can all be
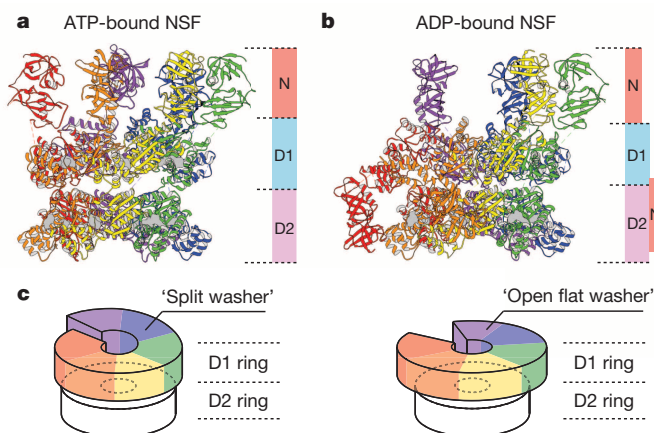
Figure 2 | Structures of ATP- and ADP-bound NSF. a, Side-view of ATP-bound NSF. **b,** Side-view of ADP-bound NSF. The six protomer chains are rainbow coloured anticlockwise based on the relative positions of the D1 domains to the D2 ring in the ATP-bound NSF model; the chain with the closest distance between D1 and D2 domains is named chain A (red). Nucleotides are shown as grey surfaces. See Methods for generation and refinement of the atomic models. **c,** A schematic diagram showing the topology of ATPase rings of ATP- and ADP-bound NSF, respectively. D1 rings are coloured according to the models shown in panels **a** and **b**.

well superposed (Extended Data Fig. 6c). Densities for ATP molecules are clearly visible in the nucleotide-binding pockets of the D1 domains of chains A through E, and in all D2 domains (Extended Data Fig. 3f, g); however, there is no clear density in the nucleotide-binding pocket of the D1 domain of chain F.

For ADP-bound NSF, the D2 ring slightly deviates from a near perfect six-fold symmetric conformation, producing a small gap (Extended Data Fig. 7a). The D1 ring is more expanded and planar compared to ATP-bound NSF (Extended Data Figs 6b and 7b), with a large opening between chains A and F, which coincides with the small gap in the D2 ring, that is, it forms an open 'flat washer' (Fig. 2c). The structures of all six D1 domains can be well superposed, but adopt different orientations relative to the D2 domains (Extended Data Fig. 6b, d).

When superposing the α/β subdomains of all the D1 domains of both ATP- and ADP-bound NSF (except for the flexible chain F in ATP-bound NSF), the α7 helix in the α subdomain is translated between the ATP and ADP-bound states (Extended Data Fig. 6e). The nucleotides are likely absent in the D1 ring of ADP-bound NSF as the conformations would not favour binding of nucleotides because of possible clashes between the nucleotide and the translated α7 helix (Extended Data Fig. 6e). This conformational change of the α subdomain is correlated with the large difference of the D1 ring between the ATP- and ADP-bound states. The ATP loaded D1 ring is more compact, with a total interface area of 5938 Å^2 compared to that of 3746 Å^2 in the ADP-bound state, resembling a spring-like transition from a 'loaded' split-washer state to a 'relaxed' open-flat-washer state (Extended Data Fig. 7b and Supplementary Video 2). This transition is further correlated with outward rotations of the D1 domains of chains A and B and the changes in their N domains (Extended Data Figs 6b and 7b, c). For a detailed comparison of our NSF structures with other members of the AAA+ family, see the Supplementary Information.

Structures of the 20S supercomplex

We prepared 20S supercomplex consisting of AMPPNP-bound hexameric NSF, αSNAP, and neuronal SNARE complex that is composed of syntaxin-1A, synaptobrevin-2/VAMP-2 (vesicle-associated membrane protein 2), and SNAP-25 (synaptosomal-associated protein 25) (Methods and Extended Data Fig. 1c–f). We used a truncated neuronal SNARE complex (green shaded fragments in Fig. 3a), identical to the one of which the high-resolution structure had been determined[12]; it was chosen because it remains monomeric in solution even at high concentration, and is sufficient for reversible assembly and disassembly (Extended Data Fig. 1g). The 3D classification of the single particle cryo-EM data of the 20S supercomplex produced four different reconstructions that each represents an asymmetric molecular state of 20S, referred to as states I, II, III$_a$ and III$_b$ (see Methods). The corresponding refined density maps of the 20S supercomplex (without symmetry) have an overall resolution ranging from 7.6 to 8.4 Å after gold-standard refinement by RELION (Extended Data Fig. 8f)[27,31].

The structures of the four states of the 20S supercomplex have the same overall architecture, so we discuss state I as a representative of all states in the following; detailed differences between the states are discussed later. The 20S supercomplex resembles a tower with different domains organized into layers (Fig. 3b). At the base of the tower (in the orientation shown in Fig. 3b, middle panel) are the D2 and D1 ATPase rings of NSF, and at the top of the tower is a 'spire', made up of four αSNAP molecules and one SNARE complex, surrounded by the six N domains of NSF. The 20S supercomplex is a striking example of broken symmetry: The approximate six-fold symmetry at the base of the complex is progressively violated in the D1 and N domains, allowing the complex to transit from six-fold symmetry to a pseudo four-fold symmetry at the top (Fig. 3b and Extended Data Fig. 8e).

The four α-helix bundle of the SNARE complex at the centre of the spire is clearly visible along with its characteristic twisted left-handed grooves, although the chemical identity of each polypeptide chain cannot be uniquely assigned at the available resolution (Fig. 3b and Extended
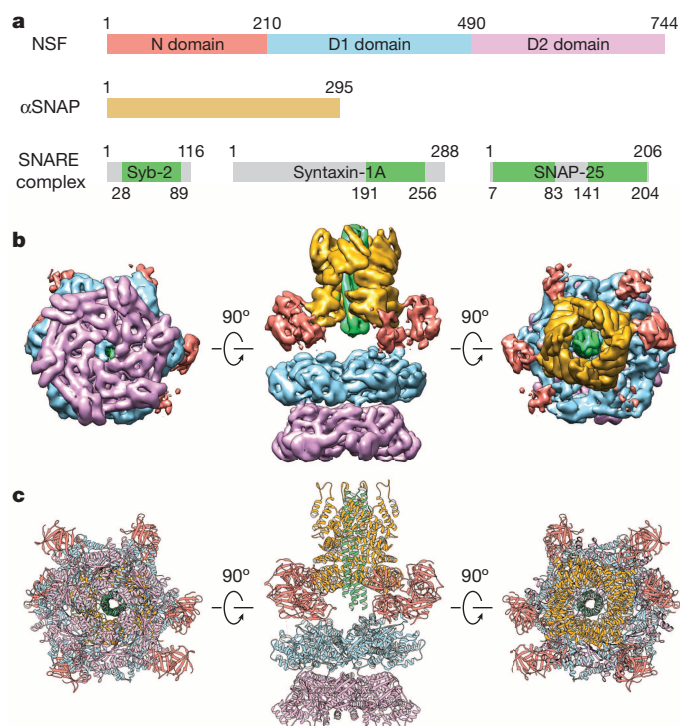
Figure 3 | 3D density map and structure of state I of the 20S supercomplex. a, Domain diagrams of the 20S supercomplex consisting of NSF, αSNAP, synaptobrevin-2 (Syb-2), syntaxin-1A, and SNAP-25. The truncated neuronal SNARE complex consists of four SNARE domains (green) from synaptobrevin-2, syntaxin-1A and SNAP-25 (two SNARE domains). The boundaries of domains and lengths of proteins are indicated above the domain diagrams. Boundaries of the fragments of the truncated neuronal SNARE complex are indicated below the diagrams. **b**, Different views of the sharpened map (4.8σ) of state I of the 20S supercomplex filtered to a resolution of 7.6 Å, with each domain colour coded to match the domain diagram in panel **a**. **c**, Three-dimensional model of state I of the 20S supercomplex. Each model is in the same orientation as in panel **b**.

Data Fig. 3l). The densities of the six N domains of NSF are not as well resolved compared to the other components of the supercomplex, but they are significantly better defined than those of ATP- and ADP-bound NSF alone. For about half of the six N domains, the density has a characteristic kidney shape as expected from the crystal structure of the N-domain (Extended Data Fig. 3j)[21,22]. To obtain an atomic model of the entire 20S supercomplex, we docked the crystal structures of the D2 and N domains[19,21], our cryo-EM structure of the D1 domain (Fig. 2a), a homology model of αSNAP derived from the crystal structure of yeast homologue Sec17 (ref. 17), and the crystal structure of the truncated neuronal SNARE complex[12] into the cryo-EM density map. Real-space rigid body minimization was carried out, followed by reciprocal space refinement as described in Methods. The resulting models fit the density well (Fig. 3c, Extended Data Fig. 3h–l, and Supplementary Video 3).

ATPase rings are tightened upon SNAP/SNARE binding

When comparing the ATPase domains of the AMPPNP-bound 20S supercomplex to those of ATP-bound NSF, there are similar overall features, along with some important differences. The overall root-mean-square-deviation (r.m.s.d.) of the main chain atoms of D1 and D2 ATPase rings is 4 Å, based on a superposition of the D1 ring. The D2 ring of 20S supercomplex is approximately six-fold symmetric, whereas the D1 ring has a split-washer-like arrangement, similar to that of ATP-bound NSF (Extended Data Fig. 7d). Overall, the ATPase rings of the 20S supercomplex adopt a tighter conformation than NSF alone (Extended Data

Fig. 7e). While there is a uniform increase in interface areas for the D2 ring, the changes are more complex for the D1 ring (Extended Data Fig. 7f). These interface area changes are correlated with conformational differences of the α subdomain of chain A and the α/β subdomain of chain F (see Extended Data Fig. 3d for definition of these subdomains), as illustrated by the relative positions of α7 helices in the two structures (Extended Data Fig. 7f, inset). Overall, the ATPase rings of NSF are tightened upon binding to αSNAPs and SNARE complex, resembling a spring being loaded (Supplementary Video 4).

Four states of the 20S supercomplex

While the D1/D2 ATPase rings of NSF are very similar among the four states, the αSNAP–SNARE spire and the N domains differ (Fig. 4a). The four states were grouped into three observed patterns (I, II, and III) based on the mode of interaction between αSNAP molecules and N domains (Fig. 4b). Considering that each αSNAP can interact with either one or two nearby N domains, one expects a total of nine theoretical patterns taking into account the split-washer asymmetry of the D1 ring (see Supplementary Discussion). However, only three patterns consistently emerged in the 3D classification. One explanation for this phenomenon is that the position of the αSNAP–SNARE spire is not random, which might favour certain patterns over the others. Indeed, the centre of the spire—the SNARE complex—is always located close to chains E and F, which are at the raised edge of the D1 split washer (arrows in Fig. 4b), suggesting possible interactions between the pore loops (YVG motif)[32] of D1 domains and the SNARE complex. For pattern III, two subclasses were refined separately, resulting in states III$_a$ and III$_b$. The main difference between the two states involves the relative position of the spires, not the pattern of αSNAP and N domain interactions (Supplementary Video 5). Thus, the 20S supercomplex exhibits four major states, which are mainly characterized by the patterns of the N domains and the position of the αSNAP–SNARE spire, whereas the conformations of the spire and base themselves do not differ much.

Multi-modal interactions between N domains and αSNAP

The structures of the four states of 20S reveal eight instances of αSNAP molecules that are interacting with two N domains of NSF, and another eight instances of αSNAP molecules that are interacting with one N domain (Fig. 4). When superposing αSNAP molecules separately based on the 1:2 and 1:1 binding scenarios, two distinct N-domain binding sites on the surface of the C-terminal region of αSNAP appear in the case of 1:2 binding, whereas in the case of 1:1 binding, the N domains

bind somewhere between these two sites (Fig. 5a). The electrostatic potential surface of the C-terminal region of αSNAP is quite negative, and both distinct binding sites are located in this negatively charged area (Fig. 5b). The N domains of NSF interact with either of the two binding sites on αSNAP via the same positively charged area (Fig. 5c). The two interfaces involve five positively charged residues of the N domain of NSF and eight negatively charged residues of the C-terminal region of αSNAP (Fig. 5d).

Previous mutagenesis studies suggested that certain positively charged residues of the N domains are important for αSNAP and SNARE binding[33], and that the C-terminal region of αSNAP is important for 20S
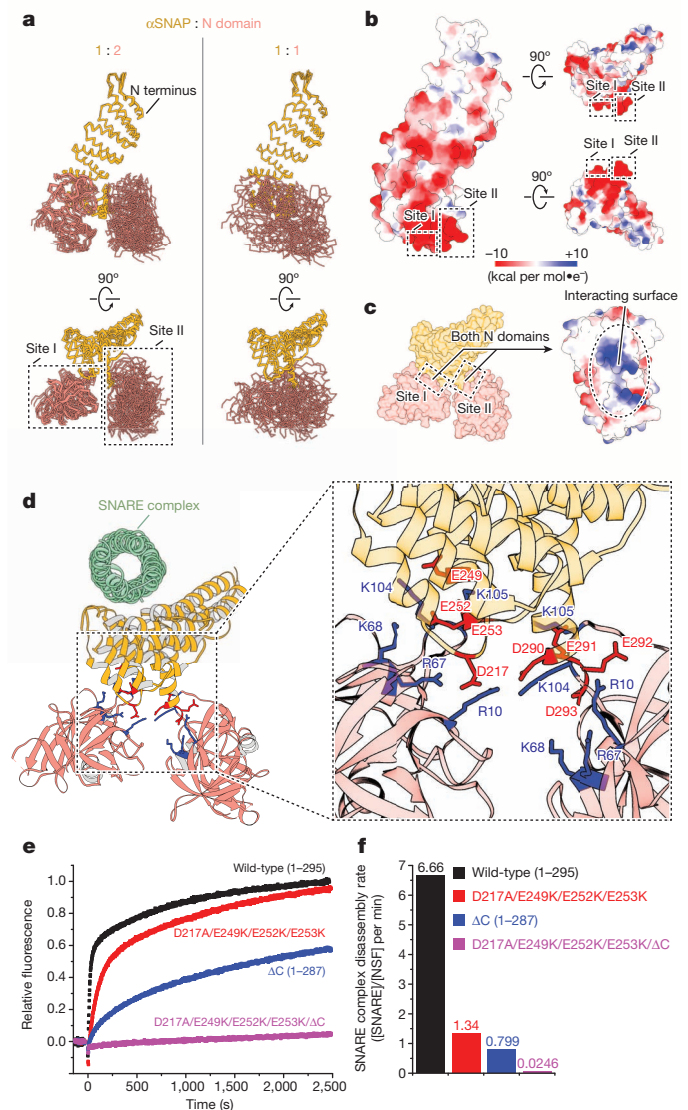
Figure 5 | Interactions between αSNAPs and N domains. a, Superposition of αSNAP molecules and N domains from the structures of the four states of 20S (Fig. 4). In eight instances one αSNAP interacts with two N domains (left). In the other eight instances one αSNAP interacts with one N domain (right). The superposition was performed with respect to the αSNAP molecules. The two distinct binding sites are highlighted by dotted boxes. **b**, Electrostatic potential surface of αSNAP. The two distinct binding sites of the N domains are highlighted. **c**, Electrostatic potential surface of the N domain. Only the surface area interacting with αSNAP molecules is shown. **d**, Ribbon representation of αSNAP and N domain interactions. The SNARE complex is shown to help with visualization. Side chains of charged residues involved in the interactions are shown, with Asp and Glu coloured in red and Arg and Lys coloured in blue. **e**, Kinetic curves of the fluorescence dequenching-based SNARE complex disassembly assay using wild-type αSNAP and specified αSNAP mutants. **f**, Corresponding initial SNARE complex disassembly rates.
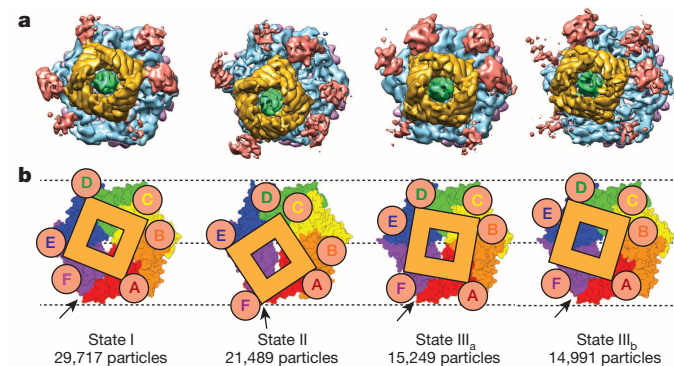
Figure 4 | Top views of the four states of 20S supercomplex. The identified four states were aligned with respect to the D1 rings. **a**, Sharpened maps (state I: 4.7σ, state II: 4.5σ, state III$_a$: 4.2σ, state III$_b$: 4.0σ). **b**, Schematic drawings to help with visualization. The N domains and αSNAP molecules are shown as spheres and squares, respectively. The D1 rings are rainbow coloured using the same scheme as in Fig. 2, with black arrows indicating the split between chain A and chain F. The D2 rings are omitted for clarity. Each N domain is labelled with its corresponding chain identifier. The numbers of particles that contributed to the reconstruction of each state are listed.

formation[34], so our structures of the 20S supercomplex now provide the molecular explanation for these previous results. To further validate our observed interfaces with a functional assay, we performed mutagenesis of αSNAP based on our 20S structures, and used a fluorescence dequenching-based assay to monitor the kinetics of SNARE complex disassembly[35]. Indeed, αSNAP mutants of either site I (quadruple mutant: D217A/E249K/E252K/E253K) or site II (C-terminal truncation: ΔC) showed impaired kinetics of SNARE complex disassembly as well as slower initial reaction rates (Fig. 5e, f). Simultaneous mutations of both sites resulted in completely inactive NSF/SNAP.

Promiscuous interaction between αSNAPs and SNAREs

In the structures of the 20S supercomplex, four αSNAP molecules wrap around the SNARE complex in a four-fold rotational symmetric arrangement, despite the fact that the SNARE complex itself is only pseudo-symmetric, that is, the four α-helices consist of different peptide chains (Fig. 6a). Even more remarkably, the αSNAP barrel has a right-handed twist in contrast to the left-handed twist of the four α-helix bundle of the SNARE complex.
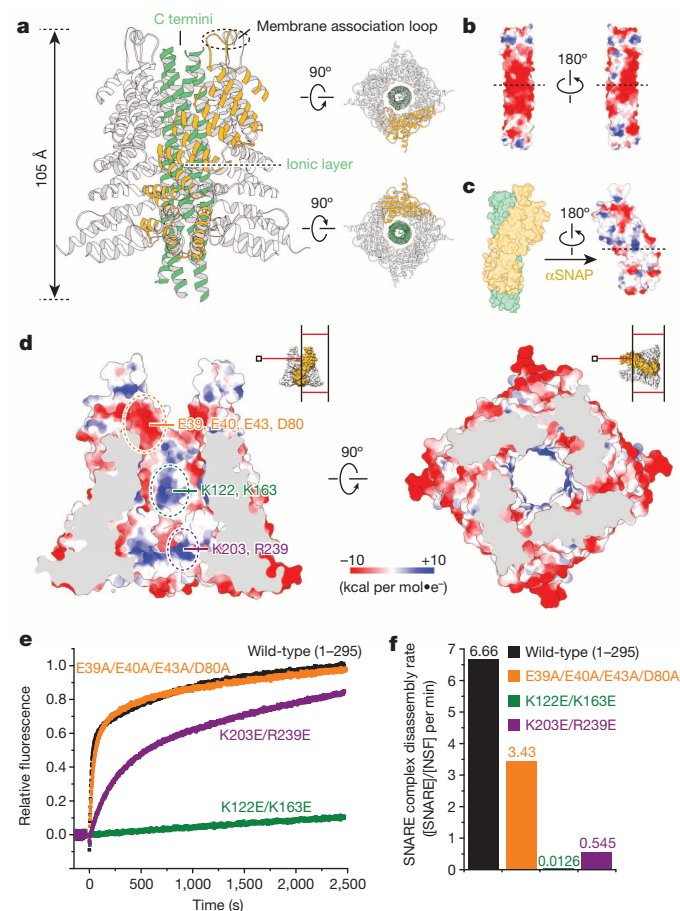
Figure 6 | Interactions between αSNAP molecules and SNARE complex.
a, A ribbon representation of the αSNAP–SNARE subcomplex for state I of 20S supercomplex. **b**, Electrostatic potential surface of the SNARE complex. Dotted lines indicate the location of the ionic layer[3,40] at the centre of the SNARE complex. **c**, Electrostatic potential surface of αSNAP. Only the surface area interacting with the SNARE complex is shown. The dotted lines indicate the location of the ionic layer of the SNARE complex. **d**, Cross-sections of the electrostatic potential surface of the αSNAP barrel. Three regions that may interact with the SNARE complex are highlighted and labelled. The locations of the cross-sections are illustrated in the insets. **e**, Kinetic curves of the fluorescence dequenching-based SNARE complex disassembly assay using wild-type αSNAP and specified αSNAP mutants. **f**, Corresponding initial SNARE complex disassembly rates.

We infer that the C terminus of the SNARE complex is facing away from NSF since it would normally continue into the transmembrane α-helices of the SNAREs syntaxin-1A and synaptobrevin-2 (transmembrane domains were not included in the constructs used). This inference is consistent with the orientations of αSNAPs, the hydrophobic loops of which are pointing away from NSF for possible membrane association[36] (Fig. 6a). Moreover, our structures provide a molecular explanation for the previous finding that the *trans*-SNARE complex before membrane fusion is resistant to NSF disassembly[37]: the αSNAP–SNARE subcomplex would not be able to form since the *trans*-SNARE complex is probably in a half-zippered state[38].

The electrostatic potential surface of the SNARE complex has a highly conserved pattern with negative charges at the centre (Fig. 6b)[39]. The interacting surface of αSNAP has two extruding positively charged residues (K122, K163) close to the central ionic layer (zero layer)[3,40] of the SNARE complex, and another two close to the C-terminal region of αSNAP (K203, R239) (Fig. 6c, d). Previous mutagenesis studies suggested the importance of K122, K163, and K203 (ref. 41). To further test the functional importance of these possible interactions, we mutated these two groups of residues, as well as a conserved concave negative patch close to the N terminus of αSNAP (E39, E40, E43, D80), and performed the SNARE complex disassembly assay. Remarkably, the K122E/K163E double mutant was completely inactive and the K203E/R239E double mutant showed impaired kinetics (Fig. 6e, f). The negative-patch quadruple mutant E38A/E40A/E43A/D80A affected kinetics to a lesser degree, with slightly decreased initial reaction rates; the presence of a membrane may make this interaction between SNAPs and SNAREs more important[36].

Variable αSNAP stoichiometry

To investigate how NSF disassembles a different SNARE complex, we prepared a SNARE complex consisting of VAMP-7 (vesicle-associated membrane protein 7), syntaxin-1A and SNAP-25 (referred to as the V7-SNARE complex); this complex includes the N-terminal Habc domain of syntaxin-1A and the N-terminal Longin domain of VAMP-7 (Fig. 7a). We assembled the 20S supercomplex with NSF and αSNAP (referred to as the V7-20S supercomplex, Extended Data Fig. 9a–c). This complex is functionally active since the V7-SNARE complex is disassembled upon ATP hydrolysis[39]. We determined a cryo-EM reconstruction to an estimated resolution of 8.0 Å without imposing any symmetry (Extended Data Fig. 9d–f). Only two αSNAP molecules bound to the rather inclined SNARE complex bundle (Fig. 7b, c). Note that the two N-terminal domains of the V7-SNARE complex and two of the six N domains of NSF were not visible. Although the spire is not as well resolved as in the 20S supercomplex (as indicated by the lack of separation of the four α-helices of the V7-SNARE complex), the cryo-EM reconstruction of V7-20S revealed that NSF can use fewer αSNAP molecules and readjust the N domains when binding to different SNARE complexes. The supercoil axis of the truncated neuronal SNARE complex in the 20S structures is approximately perpendicular to the plane of the ATPase rings, whereas in the V7-20S structure the V7-SNARE complex is angled at 76 degrees relative to the plane of ATPase rings (Fig. 7b, d). Despite these differences, the mode (for example, right-handed twist) by which αSNAP interacts with SNARE complex and the location of the SNARE complex atop the D1 ATPase ring in V7-20S are similar to those of 20S (compare Fig. 3c and Fig. 7c).

To further confirm the stoichiometry of the αSNAP–SNARE subcomplexes in solution, we conducted composition-gradient multi-angle light scattering (CG-MALS) experiments using the two different SNARE complexes mixed with αSNAP in a composition gradient (Extended Data Fig. 10a, b). The CG-MALS data analysis revealed that αSNAP binds to the truncated neuronal SNARE complex at a maximum ratio of 4:1, whereas it binds to V7-SNARE complex at a maximum ratio of 2:1 (Extended Data Fig. 10c, d). In solution, multiple species of the αSNAP–SNARE subcomplex are in equilibrium, but the cryo-EM structures of
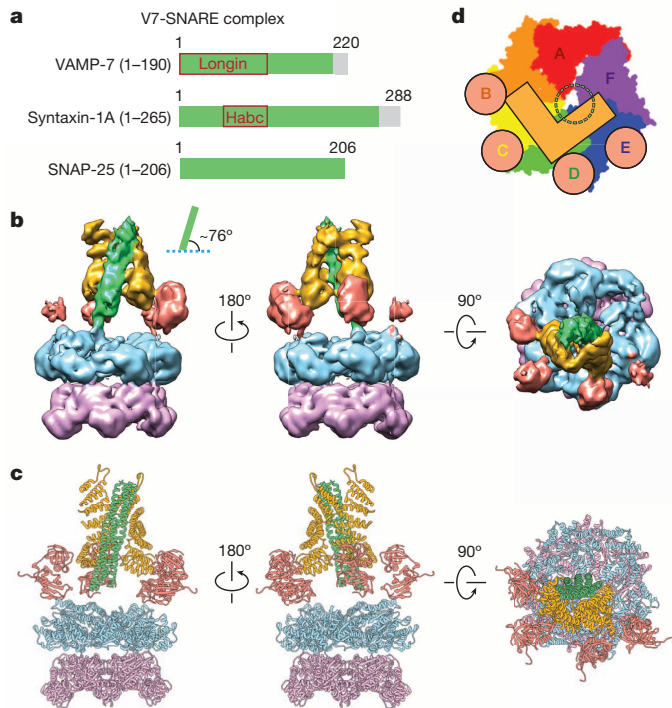
Figure 7 | 3D density map and structure of V7-20S supercomplex.
a, Domain diagram of the V7-SNARE complex. Transmembrane domains (grey) were not included in the complex. The two N-terminal domains of VAMP-7 and syntaxin-1A are highlighted by red boxes. **b**, Different views of the 3D density map (4.5σ) coloured similarly to Fig. 3b. The angle between the long axis of V7-SNARE complex and the plane of ATPase rings is approximately 76 degrees as shown in the inset. **c**, Corresponding views of the atomic model fit to the density map of V7-20S. Note that each of the two αSNAP molecules interacts with two NSF N domains. **d**, An illustration of the top view similar to Fig. 4b. The location of the N termini of the V7-SNARE α-helix bundle is indicated by a green dotted circle.

both 20S and V7-20S suggest that NSF catches the saturated complex in both cases (Extended Data Fig. 10e, f).

NSF mediated SNARE complex disassembly

Based on our cryo-EM structures, we propose a working model of NSF mediated SNARE complex disassembly (Fig. 8). Starting with the *cis*-SNARE complex (that is, with both transmembrane domains in the same membrane), the first step (Fig. 8a, b) is the binding of SNAP molecules. Our cryo-EM structures suggest that depending on the particular component proteins of the SNARE complex, up to four SNAP molecules are present. A stoichiometry higher than 4:1 is unlikely, due to packing considerations. Dozens of SNARE proteins exist in an eukaryote, but there are only a few SNAP and very few NSF species[5,18,39,42]. We propose that one NSF/SNAP species can disassemble all SNARE complexes (including yeast SNAREs[14,43]) using shape and characteristic electrostatic pattern recognition of SNARE complexes by SNAPs, rather than specific 'lock into key' interactions (Fig. 6).

The second step of our model (Fig. 8b, c) is the binding of NSF, that is, the formation of 20S supercomplex. Upon binding to the SNAP–SNARE subcomplex, which acts like a fastener, both NSF ATPase rings are tightened, akin to a loaded spring (Extended Data Fig. 7d–f and Supplementary Video 4). The N domains are immobilized due to the interactions with SNAP molecules; characteristic electrostatic patterns may also play a role in these interactions. The opposing twists of SNAP molecules and the SNARE four α-helix bundle in both the 20S and V7-20S supercomplexes (Fig. 6a), along with the existence of four distinct molecular states (Fig. 4), suggests that the 20S supercomplex exerts a torque to unwind or loosen the SNARE complex while switching between the four states. This step requires ATP hydrolysis to initiate the movement of the NSF N domains, and subsequent force transmission via SNAPs. A second possibility is that the four states represent independent binding modes: each would apply a force on its own upon binding to the SNAP–SNARE subcomplex to unwind or loosen the SNARE complex.

The final step of our model (Fig. 8c, d) is the hydrolysis of multiple ATP molecules. We observed large conformational differences of NSF between the ATP- and ADP-bound states (Fig. 1c, e and Extended Data
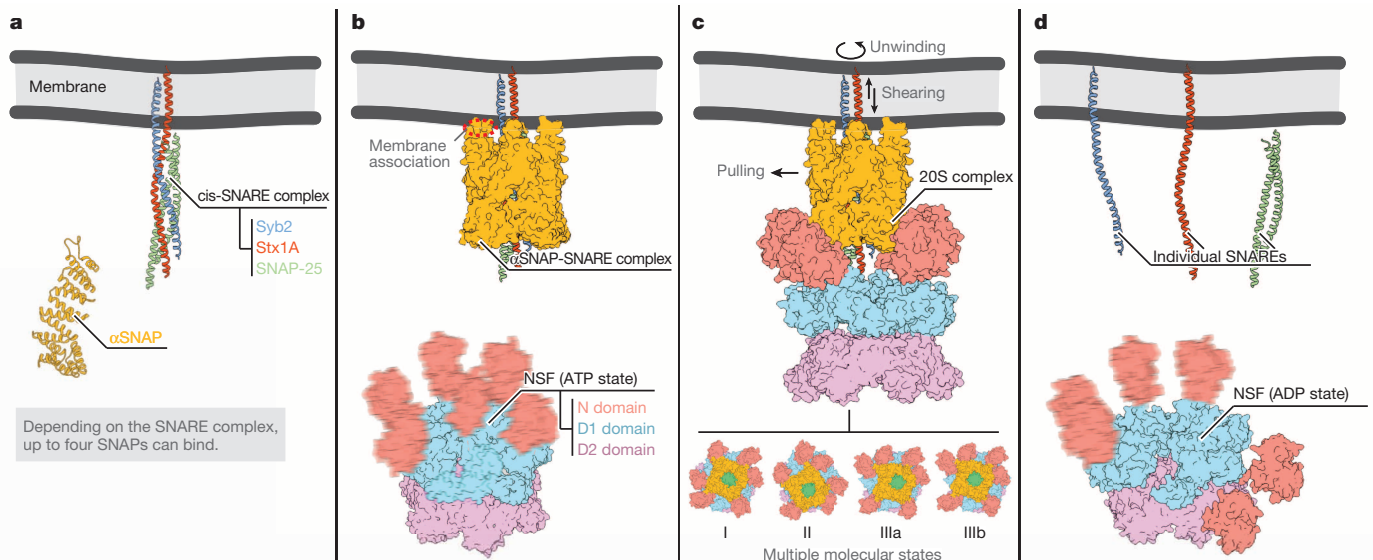
Figure 8 | Model of NSF-mediated SNARE complex disassembly. The model consists of four stages (**a**–**d**). The model refers to the neuronal SNARE complex (consisting of synaptobrevin-2 (Syb2), syntaxin-1A (Stx1A), and SNAP-25) and αSNAPs, but the model is also applicable to other SNARE complexes, along with a different number of SNAP molecules as observed in

V7-20S (see Supplementary Discussion for SNAP species and stoichiometry). The N domains and chain F of the D1 domains in panel **b** and four of the N domains in panel **d** are blurred to indicate flexibility as suggested by Cryo-EM density maps.

Fig. 7a–c), suggesting large motions of NSF upon ATP hydrolysis. The changes of the D1 ring involve a 20 Å vertical movement of the pore loops (Extended Data Fig. 6a, b), which may apply a shear force to the SNARE complex; the opening of the D1 ring from a split washer to an open flat washer, along with a motion of the N domains towards the sides of the ATPase rings, could apply a pulling force. Taken together, these forces completely disassemble SNARE complex into individual SNARE proteins. Interestingly, exactly two N domains are flipped in ADP-bound NSF (chains A and B). These two chains are at the lower edge of the D1 split washer (Fig. 2c). During the transition towards the ADP-bound flat ring, the D1 domains of chains A and B rotate outwards (Supplementary Video 4 and Extended Data Fig. 6b), so they might be in a position favoured for the motion of the N domains. Also in our 20S structure, two N domains bind to one SNAP, and together they may be able to exert a larger force compared to 1:1 binding. We speculate that state II is a precursor of ADP-bound NSF. Finally, the opening of the D1 ring in ADP-bound NSF may serve as an exit for individual SNARE proteins since a pore translocation mechanism is unlikely (see Supplementary Discussion). Our model predicts that the release of individual SNARE proteins and SNAPs would initiate nucleotide exchange and restart the cycle.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Wickner, W. & Schekman, R. Membrane fusion. *Nature Struct. Mol. Biol.* **15,** 658–664 (2008).
2. Südhof, T. C. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron* **80,** 675–690 (2013).
3. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395,** 347–353 (1998).
4. Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* **92,** 759–772 (1998).
5. Jahn, R. & Scheller, R. H. SNAREs–engines for membrane fusion. *Nature Rev. Mol. Cell Biol.* **7,** 631–643 (2006).
6. Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75,** 409–418 (1993).
7. Block, M. R., Glick, B. S., Wilcox, C. A., Wieland, F. T. & Rothman, J. E. Purification of an *N*-ethylmaleimide-sensitive protein catalyzing vesicular transport. *Proc. Natl Acad. Sci. USA* **85,** 7852–7856 (1988).
8. Malhotra, V., Orci, L., Glick, B. S., Block, M. R. & Rothman, J. E. Role of an *N*-ethylmaleimide-sensitive transport component in promoting fusion of transport vesicles with cisternae of the Golgi stack. *Cell* **54,** 221–227 (1988).
9. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 93–114 (2006).
10. Matveeva, E. A., He, P. & Whiteheart, S. *N*-Ethylmaleimide-sensitive fusion protein contains high and low affinity ATP-binding sites that are functionally distinct. *J. Biol. Chem.* **272,** 26413–26418 (1997).
11. Whiteheart, S. W., Schraw, T. & Matveeva, E. *N*-ethylmaleimide sensitive factor (NSF) structure and function. *Int. Rev. Cytol.* **207,** 71–112 (2001).
12. Ernst, J. A. & Brunger, A. T. High resolution structure, stability, and synaptotagmin binding of a truncated neuronal SNARE complex. *J. Biol. Chem.* **278,** 8630–8636 (2003).
13. Stein, A., Weber, G., Wahl, M. C. & Jahn, R. Helical extension of the neuronal SNARE complex into the membrane. *Nature* **460,** 525–528 (2009).
14. Strop, P., Kaiser, S. E., Vrljic, M. & Brunger, A. T. The structure of the yeast plasma membrane SNARE complex reveals destabilizing water-filled cavities. *J. Biol. Chem.* **283,** 1113–1119 (2008).
15. Antonin, W., Fasshauer, D., Becker, S., Jahn, R. & Schneider, T. R. Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. *Nature Struct. Biol.* **9,** 107–111 (2002).
16. Zwilling, D. *et al.* Early endosomal SNAREs form a structurally conserved SNARE complex and fuse liposomes with multiple topologies. *EMBO J.* **26,** 9–18 (2007).
17. Rice, L. M. & Brunger, A. Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. *Mol. Cell* **4,** 85–95 (1999).
18. Bitto, E. *et al.* Structure and dynamics of gamma-SNAP: insight into flexibility of proteins from the SNAP family. *Proteins* **70,** 93–104 (2008).
19. Yu, R. C., Hanson, P. I., Jahn, R. & Brunger, A. T. Structure of the ATP-dependent oligomerization domain of *N*-ethylmaleimide sensitive factor complexed with ATP. *Nature Struct. Biol.* **5,** 803–811 (1998).
20. Lenzen, C. U., Steinmann, D., Whiteheart, S. & Weis, W. Crystal structure of the hexamerization domain of *N*-ethylmaleimide–sensitive fusion protein. *Cell* **94,** 525–536 (1998).
21. Yu, R. C., Jahn, R. & Brunger, A. NSF N-terminal domain crystal structure: models of NSF function. *Mol. Cell* **4,** 97–107 (1999).
22. May, A. P., Misura, K. M., Whiteheart, S. W. & Weis, W. I. Crystal structure of the amino-terminal domain of *N*-ethylmaleimide-sensitive fusion protein. *Nature Cell Biol.* **1,** 175–182 (1999).
23. Furst, J., Sutton, R. B., Chen, J., Brunger, A. T. & Grigorieff, N. Electron cryomicroscopy structure of *N*-ethyl maleimide sensitive factor at 11 Å resolution. *EMBO J.* **22,** 4365–4374 (2003).
24. Chang, L.-F. *et al.* Structural characterization of full-length NSF and 20S particles. *Nature Struct. Mol. Biol.* **19,** 268–275 (2012).
25. Moeller, A. *et al.* Nucleotide-dependent conformational changes in the *N*-ethylmaleimide sensitive factor (NSF) and their potential role in SNARE complex disassembly. *J. Struct. Biol.* **177,** 335–343 (2012).
26. Hanson, P. I., Roth, R., Morisaki, H., Jahn, R. & Heuser, J. Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90,** 523–535 (1997).
27. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
28. DeLaBarre, B. & Brunger, A. T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nature Struct. Biol.* **10,** 856–863 (2003).
29. Huyton, T. *et al.* The crystal structure of murine p97/VCP at 3.6 Å. *J. Struct. Biol.* **144,** 337–348 (2003).
30. Davies, J. M., Brunger, A. T. & Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16,** 715–726 (2008).
31. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9,** 853–854 (2012).
32. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6,** 519–529 (2005).
33. Zhao, C., Matveeva, E. A., Ren, Q. & Whiteheart, S. W. Dissecting the *N*-ethylmaleimide-sensitive factor: required elements of the N and D1 domains. *J. Biol. Chem.* **285,** 761–772 (2010).
34. Barnard, R. J., Morgan, A. & Burgoyne, R. D. Stimulation of NSF ATPase activity by alpha-SNAP is required for SNARE complex disassembly and exocytosis. *J. Cell Biol.* **139,** 875–883 (1997).
35. Cipriano, D. J. *et al.* Processive ATP-driven substrate disassembly by the *N*-ethylmaleimide-sensitive factor (NSF) molecular machine. *J. Biol. Chem.* **288,** 23436–23445 (2013).
36. Winter, U., Chen, X. & Fasshauer, D. A conserved membrane attachment site in alpha-SNAP facilitates *N*-ethylmaleimide-sensitive factor (NSF)-driven SNARE complex disassembly. *J. Biol. Chem.* **284,** 31817–31826 (2009).
37. Weber, T. SNAREpins are functionally resistant to disruption by NSF and αSNAP. *J. Cell Biol.* **149,** 1063–1072 (2000).
38. Zorman, S. *et al.* Common intermediates and kinetics, but different energetics, in the assembly of SNARE proteins. *Elife* e03348 (2014).
39. Vivona, S. *et al.* Disassembly of all SNARE complexes by *N*-ethylmaleimide-sensitive factor (NSF) is initiated by a conserved 1:1 interaction between α-soluble NSF attachment protein (SNAP) and SNARE complex. *J. Biol. Chem.* **288,** 24984–24991 (2013).
40. Fasshauer, D., Sutton, R. B., Brunger, A. T. & Jahn, R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl Acad. Sci. USA* **95,** 15781–15786 (1998).
41. Marz, K. E., Lauer, J. M. & Hanson, P. I. Defining the SNARE complex binding surface of α-SNAP: implications for SNARE complex disassembly. *J. Biol. Chem.* **278,** 27000–27008 (2003).
42. Clary, D. O., Griff, I. C. & Rothman, J. E. SNAPs, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. *Cell* **61,** 709–721 (1990).
43. Wilson, D. W. *et al.* A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast. *Nature* **339,** 355–359 (1989).

Author Contributions M.Z., Q.Z., and D.J.C. expressed and purified all protein samples; M.Z. and S.W. collected cryo-EM data; S.W. processed cryo-EM data; M.Z. built and refined the atomic models based on cryo-EM density maps; M.Z. performed the disassembly assay; S.V. performed the CG-MALS experiment; M.Z., S.W., Y.C., and A.T.B. designed experiments, analysed data and wrote the manuscript.

Author Information 3D cryo-EM density maps of ATP- and ADP-bound NSF, and the 20S supercomplex have been deposited in the Electron Microscopy Data Bank (EMDB), with accession codes EMD-6204 (NSF-ATP), EMD-6205 (NSF-ADP), EMD-6206 (20S-state I), EMD-6207 (20S-state II), EMD-6208 (20S-state III$_a$), EMD-6209 (20S-state III$_b$) and EMD-6210 (V7–20S). The coordinates of atomic models of ATP- and ADP-bound NSF, the 20S supercomplex have been deposited in the Protein Data Bank (PDB) under the accession codes, 3J94, 3J95, 3J96, 3J97, 3J98 and 3J99. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T.B. (brunger@stanford.edu) or Y.C. (ycheng@ucsf.edu).

METHODS

Protein expression and purification. Chinese hamster NSF with a tobacco etch virus (TEV) protease cleavable N-terminal His-tag was expressed from pPROEX-1 vector in *E. coli.* BL21(DE3)-RIL cells (Agilent Technologies) at 25 °C overnight using autoinducing LB medium[44]. After collecting the cells by centrifugation, they were resuspended in lysis buffer (50 mM TrisCl, pH 8.0, 300 mM NaCl, 50 mM imidazole, and 0.5 mM TCEP), and were subjected to sonication and centrifugation. The cleared lysate was loaded onto a HisTrap column (GE Healthcare), and washed with lysis buffer. NSF was eluted using elution buffer (lysis buffer supplemented with 350 mM imidazole). The fresh elution was pooled, concentrated, and supplemented with final concentrations of 1 mM EDTA, 1 mM ATP, and 10% glycerol immediately to prevent aggregation and precipitation. The concentrated protein was immediately loaded onto a Superdex 200 16/60 column (GE Healthcare) that was pre-equilibrated with SEC Buffer (50 mM TrisCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM ATP, 1 mM DTT, and 10% glycerol). Fractions containing hexameric NSF was separated from aggregated NSF eluted from the void volume. Concentrated hexameric NSF was loaded onto a Superdex 200 16/60 column that was pre-equilibrated with monomerization buffer (50 mM sodium phosphate, pH 8.0, 150 mM NaCl, 0.5 mM TCEP). Depending on the amount of proteins, this step needed to be repeated for 3–4 times until all the NSF proteins were eluted as monomers (Extended Data Fig. 1a, b). Monomerized NSF fractions were pooled and supplemented with TEV protease, and incubated overnight at 4 °C. Tag-cleaved NSF was run through a HisTrap column to remove the TEV protease and the cleaved tags. Note that our method differs from that previously reported, which used apyrase to monomerize NSF[24]. Monomerized NSF was frozen and stored at −80 °C for future use. To reassemble the hexameric NSF in a specific nucleotide state, for example, ATP or ADP, monomerized NSF was dialysed at 4 °C overnight in reassembly buffer (50 mM TrisCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM nucleotide, 1 mM DTT, and 10% glycerol). The concentrated dialysate was loaded onto a Superdex 200 16/60 column that was pre-equilibrated with reassembly buffer for a final clearance. For sample vitrification, the final SEC buffer did not contain glycerol, and the protein samples were concentrated to ~15 mg ml^{-1}. The functional activity of purified NSF was tested by a gel-based disassembly assay (Extended Data Fig. 1g) and by a fluorescence-dequenching based SNARE complex disassembly assay described below. Note that some of the classical disassembly assays in the field were performed with NSF and nucleotide initially in the absence of Mg^{2+}, and then the reaction was triggered by the addition of Mg^{2+}, for example reference 36, establishing that NSF is active when initially prepared in complex with nucleotide in the absence of Mg^{2+}.

Rat αSNAP was expressed and purified as described before[35].

Truncated neuronal SNARE complex containing rat SNAP-25 (amino acid range 7–83), SNAP-25 (amino acid range 141–204), syntaxin-1A (amino acid range 191-256), and His-tagged synaptobrevin-2 (amino acid range 28–89) was cloned in the Duet expression system (Novagen). The complex was expressed in *E. coli.* BL21(DE3) cells at 30 °C overnight using autoinducing LB medium[44]. After collecting the cells by centrifugation, they were resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 50 mM imidazole, and 0.5 mM TCEP), and were subjected to sonication and centrifugation. The cleared lysate was loaded onto a HisTrap column (GE Healthcare), and washed with lysis buffer, urea buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 50 mM imidazole, 7.5 M urea, and 0.5 mM TCEP) and wash buffer (lysis buffer supplemented with additional 10 mM imidazole) sequentially. The SNARE complex was eluted using elution buffer (lysis buffer supplemented with 350 mM imidazole). The fresh elution was pooled and dialysed in dialysis buffer (50 mM sodium phosphate, pH 8.0, 50 mM NaCl, 3 M urea, and 0.5 mM TCEP) at 4 °C overnight for anion exchange chromatography. The anion exchange chromatography was performed in buffers containing 3 M urea (buffer A: 50 mM sodium phosphate, pH 8.0, 50 mM NaCl, 3 M urea, and 0.5 mM TCEP; buffer B: 50 mM sodium phosphate, pH 8.0, 500 mM NaCl, 3 M urea, and 0.5 mM TCEP) using a linear gradient of NaCl. The peak fractions were pooled, concentrated, and loaded onto a Superdex 75 16/60 column (GE Healthcare) that was pre-equilibrated with SEC Buffer (50 mM TrisCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP). The peak fractions were pooled and supplemented with TEV protease, and incubated overnight at 4 °C. The tag-cleaved complex was subjected to a second round of anion exchange chromatography (buffer A: 50 mM TrisCl, pH 8.0, 50 mM NaCl, and 0.5 mM TCEP; buffer B: 50 mM TrisCl, pH 8.0, 500 mM NaCl, and 0.5 mM TCEP), and further purified by a final size-exclusion chromatography in SEC buffer. The purified truncated neuronal SNARE complex was tested for disassembly activity using a SDS–PAGE gel based assay as previously described (Extended Data Fig. 1g)[35,45].

V7-SNARE complex consisting of rat full-length SNAP-25 (amino acid range 1–206), syntaxin-1A (amino acid range 1–265), and His-tagged VAMP-7 (amino acid range 1–190) was cloned and expressed similarly to the truncated neuronal SNARE complex. However, the first anion exchange chromatography step for the

truncated neuronal SNARE complex was omitted since the V7-SNARE complex showed less tendency to precipitate during purification. The other purification steps were the same as for the truncated neuronal SNARE complex.

To assemble the 20S supercomplex, hexameric NSF loaded with AMPPNP was mixed with αSNAP and truncated neuronal SNARE complex in a mole ratio of 1:10:2, and incubated on ice for 30 min. The mixture was then concentrated and purified by size-exclusion chromatography using a Superdex 200 10/300 column (GE Healthcare) pre-equilibrated with 20S Buffer (50 mM TrisCl, pH 8.0, 150 mM NaCl, 1 mM AMPPNP, 1 mM EDTA, and 1 mM DTT) (Extended Data Fig. 1c–f). The resulting peak fractions containing the 20S supercomplex were pooled and concentrated to a final concentration of ~15 mg ml^{-1} for cryo-EM studies. The assembly and purification protocol for V7-20S supercomplex were essentially the same, except that hexameric NSF loaded with ATP was used.

Sample vitrification. Initial attempts to image the reassembled NSF and 20S supercomplex by cryo-EM were hindered by preferential orientations. 2D class averages showed that most of the particles were in end-on views, which is insufficient for structure determination (Extended Data Fig. 8a, inset). When the samples were deposited to a thin layer of carbon on top of the holey carbon grid, many side views were observed under FEI Tecnai TF20 operated at 200 kV. However, the contrast of the particles was significantly weaker. To ensure a sufficient number of side-view particles suspended in vitreous ice, samples were incubated in a buffer containing a small amount of detergent before plunge freezing. Multiple detergents were screened, and particles in Nonidet P-40 or its substitutes displayed the most side views. The final protein solution contained 0.05% of Nonidet P-40. However, this buffer condition dramatically reduces the particle density in the hole (~100-fold). In order to achieve a reasonable particle distribution, we concentrated the protein sample to approximately 15 mg ml^{-1}. This concentration exceeds the usual requirement for cryo-EM. However, even at such high concentration, we observed few particles in areas where ice thickness is considered to be thin and ideal for cryo-EM. Therefore, images were collected from holes with ice that was considered to be thick by general cryo-EM consensus. Quantifoil Cu R1.2/1.3 grids (Quantifoil Micro Tools GmbH, Germany) were washed in chloroform for one hour and air-dried overnight. Aliquots of 2.5 μl samples were loaded onto the grids. Because the buffer contained detergent, the protein solution spread relatively well over the grid surface without glow discharge. Grids were blotted for 3 to 4 s and plunge frozen in liquid ethane cooled by liquid nitrogen using a FEI Vitrobot (FEI Company). The same vitrification conditions were applied to all samples (ATP- and ADP-bound NSF, 20S and V7-20S).

Cryo-EM data collection. Grids were transferred to TF30 Polara equipped with a field emission source and operated at 300 kV. Images were recorded on a Gatan K2 Summit direct electron detector operated in super-resolution counting mode following the established dose fractionation data acquisition protocol[46]. Images were recorded at a nominal magnification of 31000×, corresponding to a calibrated super resolution pixel size of 0.61 Å on the specimen. The dose rate on the detector was set to be ~8.2 counts (corresponding to ~10.9 electrons) per physical pixel per second. At this setting, the coincidence loss is about 11.5% and the total loss, including the loss due to imperfect detector quantum efficiency, is about 25%[47]. The total exposure time was 6 s, leading to a total accumulated dose of 44 e$^-$ per Å^2 on the specimen. Each image was fractionated into 30 frames, each with an accumulation time of 0.2 s. Dose-fractionated images were recorded using a semi-automated acquisition program UCSFImage4 (written by Xueming Li). Defocus values ranged from −1.8 to −2.8 μm.

Image processing. Super-resolution counting images were 2 × 2 binned by Fourier cropping for motion correction, resulting in a pixel size of 1.22 Å. Motion corrected frames were summed to a single micrograph for subsequent processing[46]. Defocus values were determined for each micrograph using CTFFIND3[48]. A semi-automated procedure similar to a previous work was used to pick particles[49]. Briefly, for each data set, ~2,000 particles were manually picked to calculate 2D class averages. Unique 2D class averages were selected as templates for automated particle picking. Picked particles were visually inspected. Obviously falsely picked particles were removed. 2D classification was performed using RELION and SPIDER[50]. Initial 3D models were generated using the common lines method[51]. 3D classification and gold-standard refinement were performed using RELION[27]. The initial model from the common lines method was low-pass filtered to 60 Å and used as the starting model for the initial auto-refinement, which generated a consensus model. This consensus model was again filtered to 60 Å and used for 3D classification. We used prior knowledge of NSF to distinguish 'good' classes from 'bad' classes: 3D class averages that showed incorrect features of NSF were considered as bad, including the wrong numbers of apparent D1 and/or D2 domains or a seriously deteriorated D2 ring. All good class averages had the correct number (six) of all the domains and well-defined D2 ring density.

For 20S, good 3D class averages showed the correct number of all the domains, and well-defined densities of the αSNAP–SNARE spire and D2 ATPase ring. Particles

in distinct good 3D classes were refined separately, yielding four final reconstructions (Fig. 4a). The four final reconstructions belong to three patterns (I, II, and III) based on the mode of interactions between αSNAP and N domains. When we tested different settings of the 3D classification procedure[27], other patterns were either rarely populated or showed deteriorated features in some domains, preventing refinement of the pattern to a reasonable resolution. Therefore, these other patterns may represent minor populations if they actually exist. Note that when we sub-classified pattern I or II, we also observed multiple states, but the differences between them were much smaller than those between states III$_a$ and III$_b$. Therefore we did not refine those subclasses individually. The observed conformational heterogeneity may have prevented crystallization of the 20S supercomplex despite extensive efforts by many laboratories.

No symmetry was assumed throughout the entire process for all NSF and 20S reconstructions (except for demonstrating the detrimental effect of including symmetry, Extended Data Fig. 5, see also Supplementary Discussion). The solvent area of raw particles was set at zero for 3D classification (zero_mask = true). The solvent area was filled with random noise for auto-refinement (zero_mask = false). The result from auto-refinement was slightly worse when zero_mask was set to true. For ATP-bound NSF sample, continued refinement using the same particles but summed from frame number 2 to number 18 (the first frame is number 1) improved the resolution from 4.4 Å to 4.2 Å. As recommended by the RELION documentation, "relion_postprocess" was used to generate a soft mask, which was applied to the two half maps before FSC was calculated. The resolution reported in Extended Data Table 1 was estimated from the masking-effect-corrected FSC curve using the FSC = 0.143 criterion[52]. B-factor sharpening was carried out by "relion_postprocess" implementing an automated B-factor fitting algorithm[53]. The local resolution was estimated using ResMap on unsharpened and unfiltered maps[54]. Detailed information for each final reconstruction is summarized in Extended Data Table 1.

Model building and refinement. The two half-maps of ATP-bound NSF were summed to a single unsharpened map. The summed map was then sharpened manually using a series of negative B-factors by XMIPP[55]. The sharpened maps were used for model building in COOT[56]. The crystal structure of the D2 domain of NSF (PDB accession code: 1NSF)[19] in complex with ATP-Mg^{2+} was first docked as the hexamer observed in the crystal structure into the density map as a single rigid body. Minor adjustments of the backbone and side chains of the D2 domain were manually performed using COOT. The nucleotide conformation observed in the crystal structure matched the density of the cryo-EM map (Extended Data Fig. 3f), demonstrating that the absence of Mg^{2+} in the assembly of ATP-bound NSF did not alter the conformation of the nucleotide. A homology model of the D1 domain protomer generated by SWISS-MODEL[57] based on the crystal structure of N-D1 domain of p97 (PDB accession code: 1E32)[58] was initially placed into the best of the six copies of D1 densities. However, the detailed fit to the observed densities of the D1 domains was relatively poor and it required complete retracing the main chain and side chains in order to obtain a better fit (starting from the structure of the D2 domain of p97 as the homology model did not produce a better initial fit to the density map and it would have still required complete retracing). The rebuilt D1 domain was then docked into the other five copies of D1 density. Densities for ATP were visible in 11 out of the 12 possible nucleotide-binding pockets (no clear density for nucleotide was seen in the D1 domain of chain F), and models of ATP were fit to these densities. Rigid body minimization was carried out for the α/β and α subdomains of each D1 domain separately using COOT, followed by manual adjustment, residue by residue. To complete the model, the crystal structure of NSF N domain (PDB accession code: 1QCS)[21] was docked into the corresponding densities of the unsharpened map without any fitting; the quality of these densities was good enough to determine the approximate positions of the N domains. The partial model containing the D1 and D2 domains was subjected to reciprocal space refinement. Amplitudes of the summed map were corrected by frequency-dependent scaling factors determined by comparing the experimental maps with a reference map calculated from the full model[59,60]. A soft-edged mask was generated based on the built atomic model (including only the D1 and D2 domains) and applied to the scaled map. Most solvent regions and the density corresponding to N domains were masked out. The masked maps were put into an artificial unit cell with P1 symmetry and converted to MTZ format using CCP4 program sftools[61]. The resulting reflection files were used to perform maximum likelihood refinement using PHENIX[62] with secondary structure restraints, reference model restraints, and automatic optimization of data and stereochemistry weights. The reference models were generated from the built models using the geometry minimization function in PHENIX. The refined D1 and D2 domains were combined with the docked N domains to produce a complete model.

To model ADP-bound NSF, the crystal structure of the D2 domain and the refined cryo-EM structure of the D1 domain (from ATP-bound NSF) were docked into the sharpened map. Rigid body minimization was carried out for the α/β and α subdomains of each ATPase domain (six D1 and six D2) separately using COOT.

Densities for ADP were clearly visible for four nucleotide-binding sites of the D2 domains (except those in chains A and F). The resulting model was refined using PHENIX[62] as described above for the ATP-bound structure. The model was completed by docking the N domains into the unsharpened map, without fitting and further refinement; the quality of these densities was good enough to determine the approximate positions of the N domains.

For 20S supercomplex structure, the model of ATP-bound NSF containing the D1 and D2 domains was docked into each of the four sharpened maps corresponding to the four states of 20S (Fig. 4). Rigid body minimization was carried out for the α/β and α subdomains of each ATPase domain (six D1 and six D2) separately using COOT[56]. The crystal structure of the N domain (PDB accession code: 1QCS)[21] was docked into the peripheral densities. The SNARE complex was easily identified and modelled using the crystal structure of truncated neuronal SNARE complex (PDB accession code: 1N7S)[12]. A homology model of αSNAP generated by SWISS-MODEL[57] based on the crystal structure of yeast Sec17 (PDB accession code: 1QQE)[17] was docked into the remaining densities surrounding the SNARE complex. The resulting models were refined in reciprocal space using PHENIX[62] as described above.

The fitting of the V7-20S supercomplex was performed similarly, except that only two αSNAPs and four out of six N domains were modelled based on the observed densities in the cryo-EM map. The crystal structure of the closely related truncated neuronal SNARE complex[12] was used for model building since there is no crystal structure of the V7-SNARE complex available. The Habc domain of syntaxin-1A and the Longin domain of VAMP-7 were not visible. The model of V7-20S supercomplex was not further refined in reciprocal space.

For cross-validation of overfitting, we followed the procedures published before[63]. Briefly, the coordinates of refined ATP-bound NSF model (only D1 and D2 domains) were displaced randomly by 0.1 Å using PHENIX (PDB Tools) to remove potential model bias. The displaced model was then refined against one of the half maps in reciprocal space. FSC curves were calculated between the resulting model and half map 1 ('work', that is, used for refinement), the resulting model and half map 2 ('free', that is, not used for refinement), and the resulting model and the summed map. The lack of separation between work and free FSC curves suggested that the model was not overfitted.

MolProbity[64] was used for evaluating the geometry of the models. The model statistics of the refined models are summarized in Extended Data Table 2 (note that R values are arbitrary since they depend on the exact definition of the P1 cell, so they are not provided in the table). Molecular graphics and analyses were performed with either PyMOL (The PyMOL Molecular Graphics System, Version 1.7.0.5 Schrödinger, LLC.) or UCSF Chimera[65] (Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, supported by NIGMS P41-GM103311). All density maps presented were B-factor sharpened and filtered by RELION using the B-factors listed in Extended Data Table 1, unless otherwise mentioned.

Flourescence dequenching-based SNARE complex disassembly assay. The details of the assay were published previously[35]. Briefly, soluble rat neuronal SNARE complex containing syntaxin-1A (amino acid range 1–265, S249C, K253C), full-length SNAP-25 (amino acid range 1–206), and synaptobrevin-2 (amino acid range 1–96), was labelled with Oregon Green (forming covalent linkages with two cysteine residues close to the C terminus of syntaxin-1A, and four native cysteine residues of SNAP-25). The disassembly reactions were carried out using FlexStation II (Molecular Devices) in a 384-well plate with a reaction volume of 60 μl. Each condition (different αSNAP mutants) was divided into four replicas, and the average was plotted. Final concentrations of 400 nM Oregon Green-labelled SNARE complex, 2 μM αSNAP, and 85 nM NSF were included in the reaction buffered with 50 mM TrisCl, pH 8.0, 20 mM NaCl, 2 mM ATP, 2 mM MgCl$_2$, and 0.5 mM TCEP. To measure the initial SNARE complex disassembly rate, the first 20 data points after adding the NSF were used for linear regression analysis, except for αSNAP mutants D217A/E249K/E252K/E253K/ΔC, and K122E/K163E, for which the first 1,000 data points were used because the slope was close to zero. Note that the disassembly rate of wild type proteins (Fig. 5f) is comparable to that of previous experiments[35,39], illustrating that our purification method of NSF produces fully active proteins.

Composition-gradient multi-angle light scattering (CG-MALS). The experiments were performed in Tris buffered saline (TBS; 17 mM Tris, 50 mM NaCl, pH 8.0) with or without 0.1 mM TCEP. After dilution to the appropriate stock concentration, proteins were filtered using a syringe-top filter (0.02 μm, Anotop, Whatman). CG-MALS experiments were performed with a Calypso II composition gradient system (Wyatt Technology Corporation) to prepare different compositions of protein and buffer, and deliver them to an online UV/Vis detector (Waters Corporation) and DAWN HELEOS II multi-angle light scattering detector (Wyatt). Inline filter membranes with 0.03 μm pore size were installed in the Calypso system for additional sample and buffer filtration. An automated CG-MALS method was performed, consisting of single component concentration gradients for each species to quantify

any self-association and a dual-component 'crossover' gradient to assess the interaction between αSNAP and each SNARE complex (Extended Data Fig. 10a). For each composition, the Calypso system prepared an aliquot of protein solution, injected it into the detectors and stopped the flow for 60–240 s to allow the reaction to come to equilibrium within the MALS detector flow cell. Equilibrium light scattering and concentration data were fit to an appropriate association model using the CALYPSO software (Wyatt).

44. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
45. Fasshauer, D., Eliason, W. K., Brunger, A. T. & Jahn, R. Identification of a minimal core of the synaptic SNARE complex sufficient for reversible assembly and disassembly. *Biochemistry* **37,** 10354–10362 (1998).
46. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).
47. Li, X., Zheng, S. Q., Egami, K., Agard, D. A. & Cheng, Y. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J. Struct. Biol.* **184,** 251–260 (2013).
48. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
49. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504,** 107–112 (2013).
50. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
51. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3,** 1941–1974 (2008).
52. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135,** 24–35 (2013).
53. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11,** 63–65 (2014).
55. Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nature Protocols* **3,** 977–990 (2008).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
57. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–W258 (2014).
58. Zhang, X., Shaw, A., Bates, P. & Newman, R. Structure of the AAA ATPase p97. *Mol. Cell* **6,** 1473–1484 (2000).
59. Penczek, P., Ban, N., Grassucci, R. A., Agrawal, R. K. & Frank, J. Haloarcula marismortui 50S subunit-complementarity of electron microscopy and X-Ray crystallographic information. *J. Struct. Biol.* **128,** 44–50 (1999).
60. Saad, A. *et al.* Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J. Struct. Biol.* **133,** 32–42 (2001).
61. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).
62. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
63. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343,** 1485–1489 (2014).
64. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
65. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
66. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
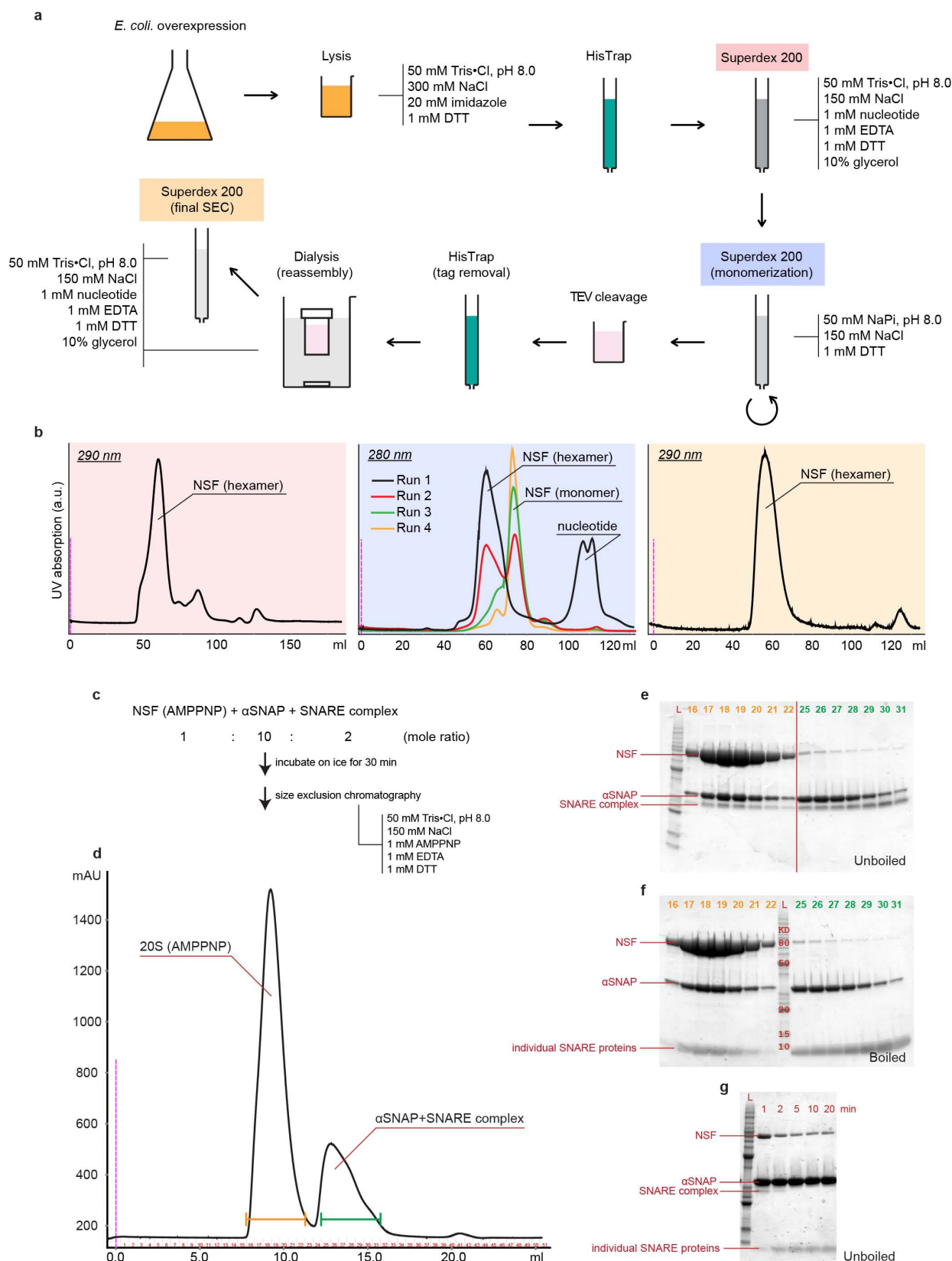
Extended Data Figure 1 | Purification of recombinant NSF bound to specific nucleotide and 20S supercomplex. **a**, Schematic diagram of the purification steps of NSF. Chromatography columns and buffer conditions are provided. **b**, Size-exclusion chromatograms corresponding to the coloured steps in panel **a**. Major peaks are labelled. **c**, A scheme showing the purification steps of the 20S supercomplex. **d**, Size-exclusion chromatogram of the 20S supercomplex. Major peaks are labelled. **e**, SDS–PAGE gel of fractions collected in panel **d**. The samples were not boiled. **f**, SDS–PAGE gel of the same fractions as in panel **e**. The samples were boiled. **g**, SDS–PAGE-based SNARE disassembly assay of the truncated neuronal SNARE complex. This complex is stable in SDS without boiling. Disassembly by NSF/αSNAP is observed as a function of time.
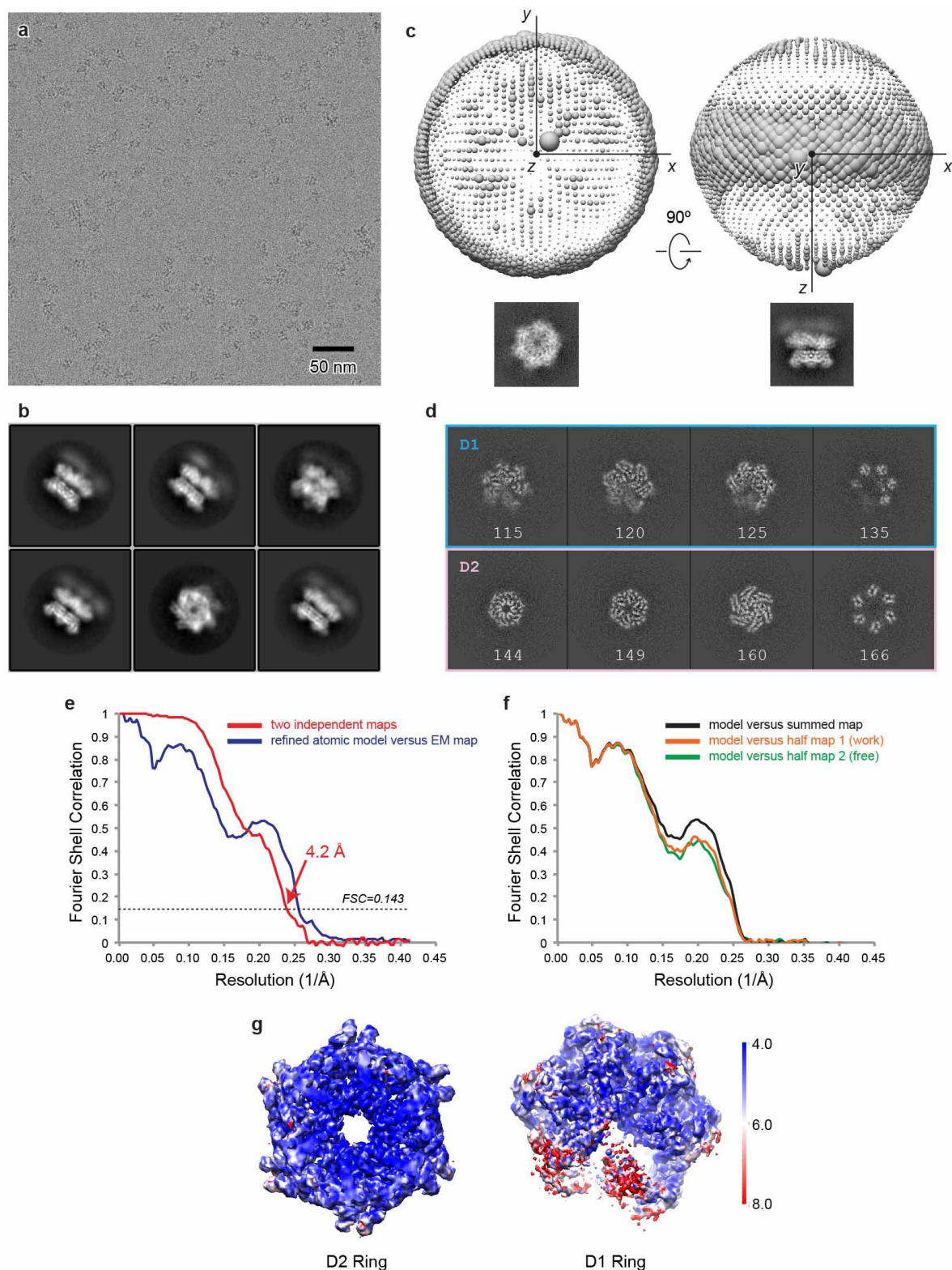
Extended Data Figure 2 | 3D reconstruction of ATP-bound NSF by single-particle cryo-EM. a, A representative electron micrograph (out of 1,150 micrographs) of ATP-bound NSF particles in vitreous ice. **b**, Selected 2D class averages (6 out of 50). **c**, Plots of the angular distribution of particle projections. The radius of the sphere at each projection direction is proportional to the number of particle images assigned to it. Two alternative views are shown, with either the *z* axis or the *y* axis pointing out towards the viewer. Two corresponding re-projection images of the final density map are shown under the plots. **d**, Selected slice views of the final reconstruction. The slice numbers are indicated. **e**, FSC curves for the 3D density map after RELION post-processing. The resolution is estimated to be 4.2 Å by the gold-standard refinement criterion, as indicated by the red arrow. The FSC curve between the refined atomic model and the 3D density map is shown in blue. **f**, FSC curves for cross-validation. Black, model versus summed map (full data set); green, model versus half map 1 (used for test refinement); orange, model versus half map 2 (not used for test refinement). See Methods for details. **g**, 3D density map coloured according to the local resolution as estimated by ResMap.

a

Chain C

Y593

α3 (D2)

F589

A372

I370

L368

3.8 Å

T373

3.8 Å

V369

β4 (D1)

I367

b

Chain C (D1)

β2

β3

β4

β1

β5

c

D1

linker

D2

Chain B

d

α subdomain   ATP   α/β subdomain

pore loop

α7 helix   α2 helix

Chain C (D1)

e

Chain C (D1)

ATP

Sensor I

Walker B

Walker A

Sensor II

Arginine finger (chain B)

f

Chain C (D2)
superposed with crystal structure 1nsf

Magnesium

ATP

g

Chain A (D1)   *Chain B (D1)*   *Chain C (D1)*

Chain D (D1)   *Chain E (D1)*   *Chain F (D1)*

h   i   j   k   l

D2 ring   D1 ring   N domain   αSNAP   SNARE complex
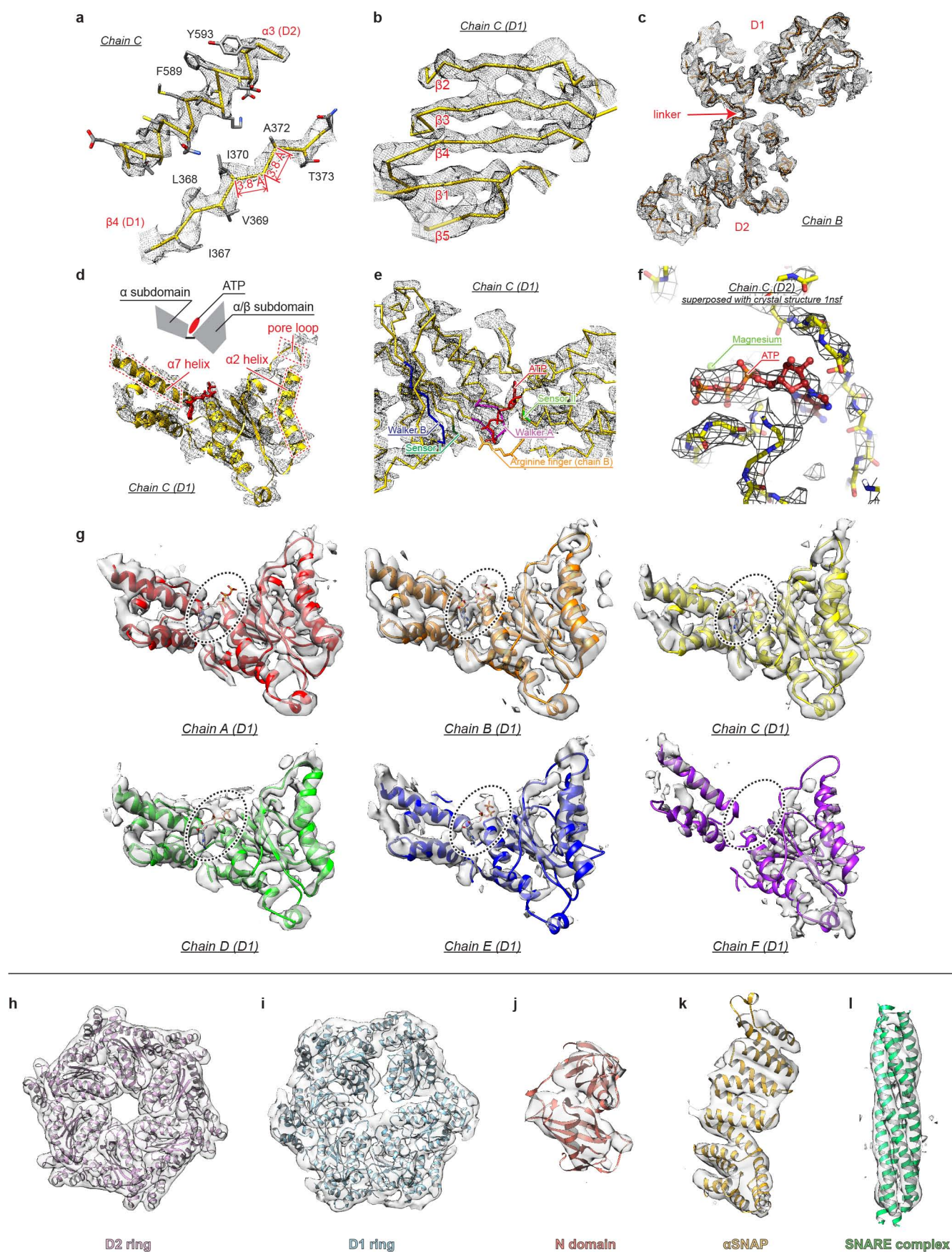
Extended Data Figure 3 | Representative densities from the cryo-EM reconstructions of ATP-bound NSF and the 20S supercomplex. **a**–**g**, ATP-bound NSF. Density maps were sharpened by XMIPP using a B-factor of $-123\,\text{Å}^2$. **a**, Representative densities (black mesh, 7.8σ) for an α-helix and a β-strand of the D1 domain with the refined model (coloured sticks) superposed. **b**, Representative density (black mesh, 7.0σ) of the β-sheet of the α/β subdomain of D1 (chain C) with the refined model (yellow Cα ribbon) superposed. **c**, Density (black mesh, 7.0σ) and model (yellow Cα ribbon) for the D1 and D2 domains of chain B. Note that the linker between the two domains is well resolved. **d**, *De novo* model (yellow cartoon) of the D1 domain built from the cryo-EM density map (black mesh, 7.0σ). The arrangement of the subdomains and nucleotide is illustrated in the inset. The pore loop (YVG motif) and two α-helices: α2 from the α/β subdomain, and α7 from the α subdomain are highlighted in the red dotted boxes. **e**, Density (black mesh, 6.5σ) and model (yellow Cα ribbon) of the ATP binding pocket of the D1 domain (chain C). Motifs that are typical for AAA+ ATPases are indicated.

f, Superposition of the crystal structure of the ATP-bound D2 domain with Mg^{2+} (ref. 19, PDB accession code: 1NSF, coloured sticks and balls), and the cryo-EM density map (black mesh, 7.6σ) of ATP-bound NSF (density of chain C). The crystal structure was docked into the density as a rigid body without any refinement. Note that no Mg^{2+} was present in the samples for cryo-EM studies, but the ATP molecule and the protein coordinates from the crystal structure match the cryo-EM density well. **g**, Nucleotide-binding sites of the D1 domains from ATP-bound NSF. The density (translucent surface, chains A–E: 8.2σ, chain F: 5.0σ) of each D1 domain is shown together with the built model in ribbon representation. The nucleotide-binding pockets are highlighted by dotted circles. Five out of the six D1 domains show clear density for ATP. **h**–**l**, Representative densities (translucent surface, 4.7σ) from the reconstruction for state I of the 20S supercomplex with the model (cartoon) superposed. All densities are representative except for the N domain in panel **j**, which represents the better-resolved half of the N domain densities (12 out of 24 cases).
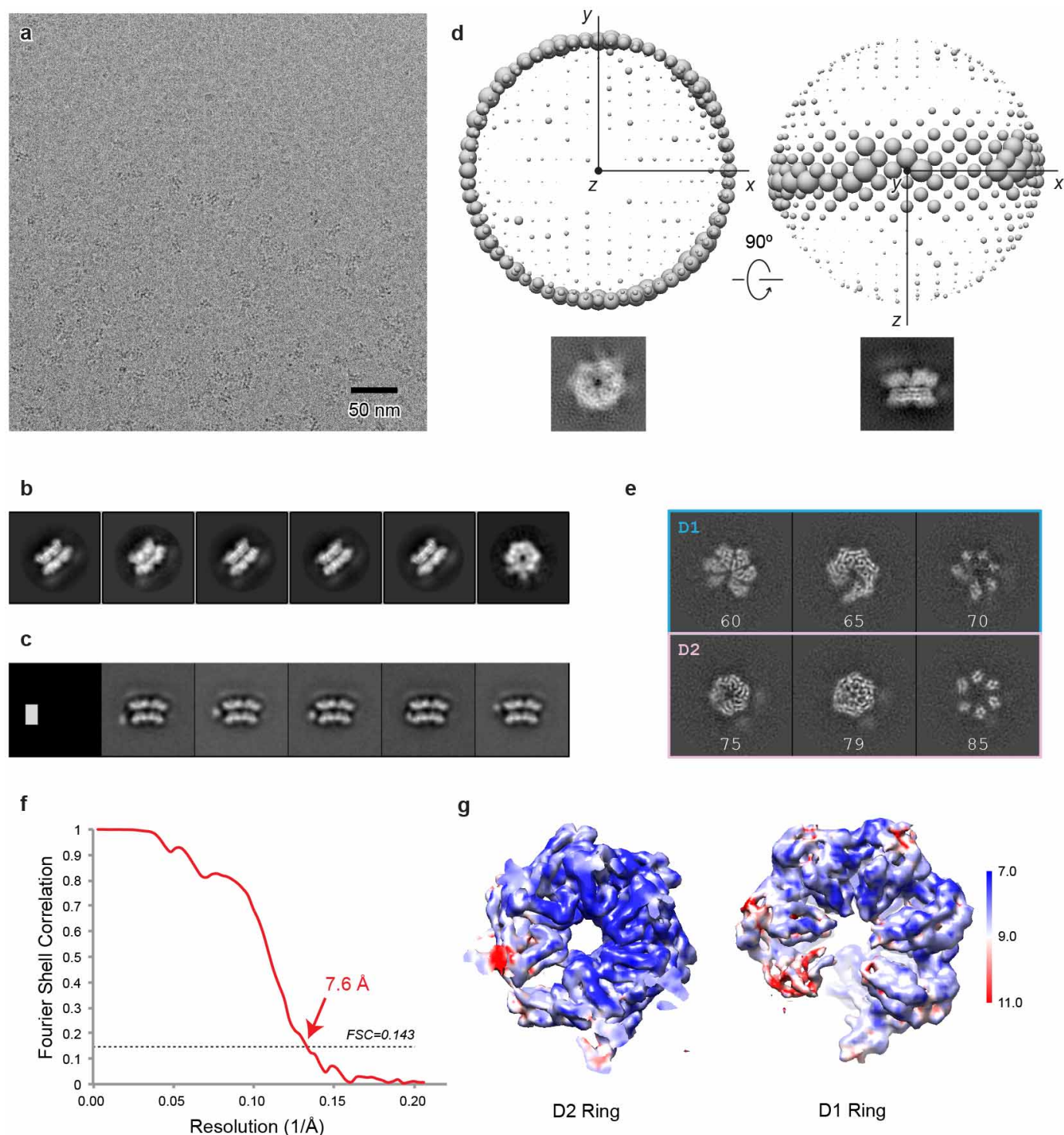
Extended Data Figure 4 | 3D reconstruction of ADP-bound NSF by single-particle cryo-EM. a, A representative electron micrograph (out of 840 micrographs) of ADP-bound NSF particles in vitreous ice. **b**, Selected 2D class averages (6 out of 30). **c**, Selected focused 2D class averages (5 out of 10). The first image shows the focused classification mask, which locates the flipped N domains. **d**, Plots of angular distribution of particle projections. The radius of the sphere at each projection direction is proportional to the number of particle images assigned to it. Two alternative views are shown, with either the *z* axis or the *y* axis pointing out towards the viewer. Two corresponding re-projection images of the final density map are shown under the plots. **e**, Selected slice views of the final reconstruction. Slice numbers are indicated. **f**, FSC curve for the 3D density map after RELION post-processing. The resolution is estimated to be 7.6 Å by the gold-standard refinement criterion. **g**, 3D density map coloured according to the local resolution as estimated by ResMap.
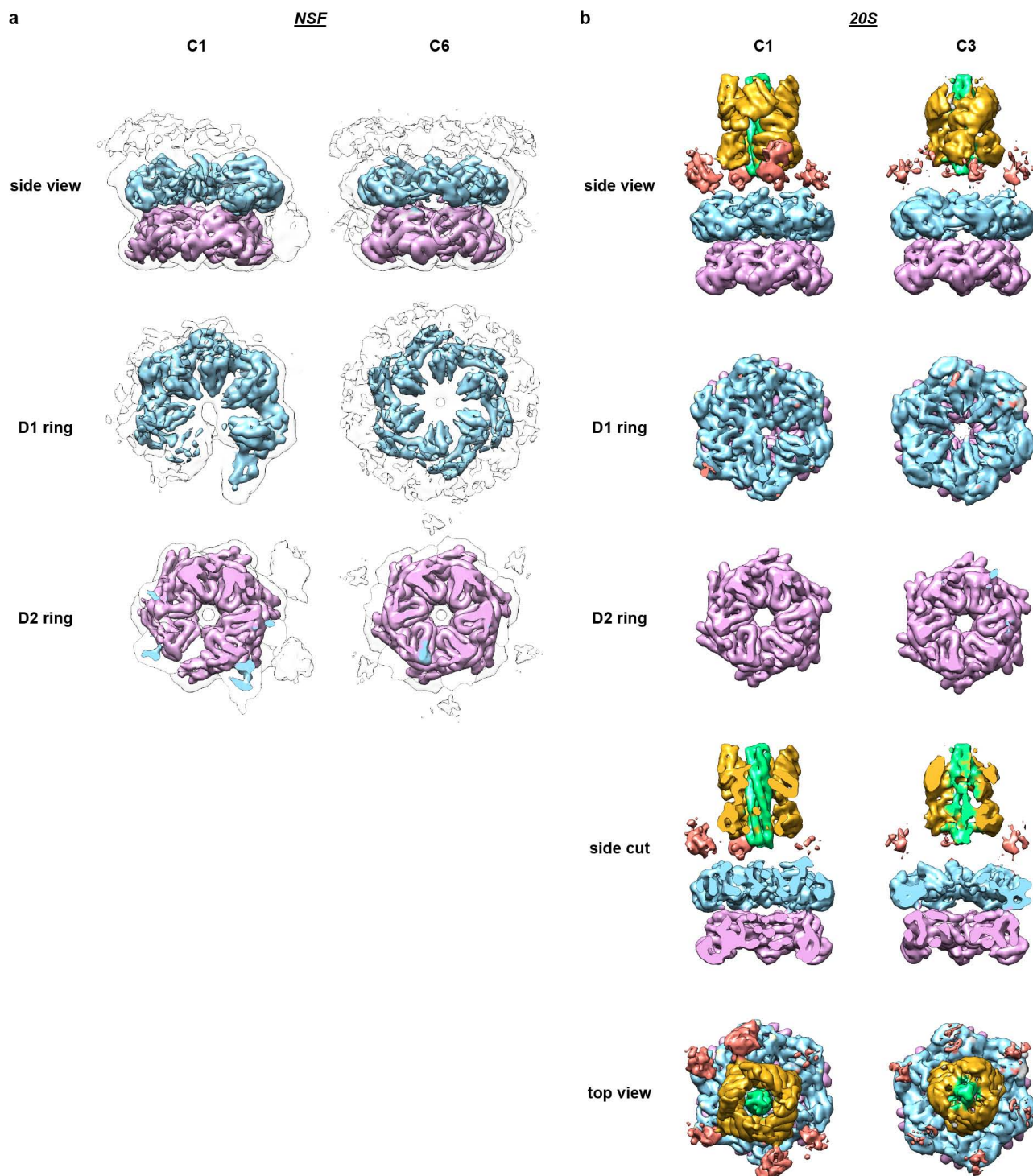
Extended Data Figure 5 | Detrimental effect of imposing C6 symmetry on the reconstruction of ADP-bound NSF and C3 symmetry on the reconstruction of the 20S supercomplex. **a**, For the NSF maps, in order to visualize densities of the N domains, an unsharpened map is displayed (translucent surface, C1: 1.2σ, C6: 0.6σ) together with the sharpened map using no symmetry (C1) or C6 symmetry during reconstruction (coloured surface, C1: 5.9σ, C6: 7.0σ). For the reconstruction that uses C6 symmetry, symmetric densities for the N domains at top and side positions appear in the unsharpened map, however, these densities cannot be matched to the crystal structure of the N domain. Likewise, the D1 domains appear compressed and cannot be fit well using the structure of the D1 domain that we obtained by asymmetric reconstruction. **b**, Reconstruction of state I of the 20S supercomplex without symmetry (C1) or with C3 symmetry. Density maps are shown in coloured surfaces similar to Fig. 3 (C1: 4.7σ, C3: 4.9σ). The C3 averaging causes the D1 domains to display alternating up and down positions. The density for the SNARE complex is a featureless rod without the characteristic left-handed twist of the four α-helix bundle. Densities for only three SNAPs emerge, but without any interpretable features (for example, there are no grooves between helices), preventing a match with the crystal structure of the known homologue of αSNAP, Sec17. The N domain densities are weak and none of them exhibit the expected kidney shape.
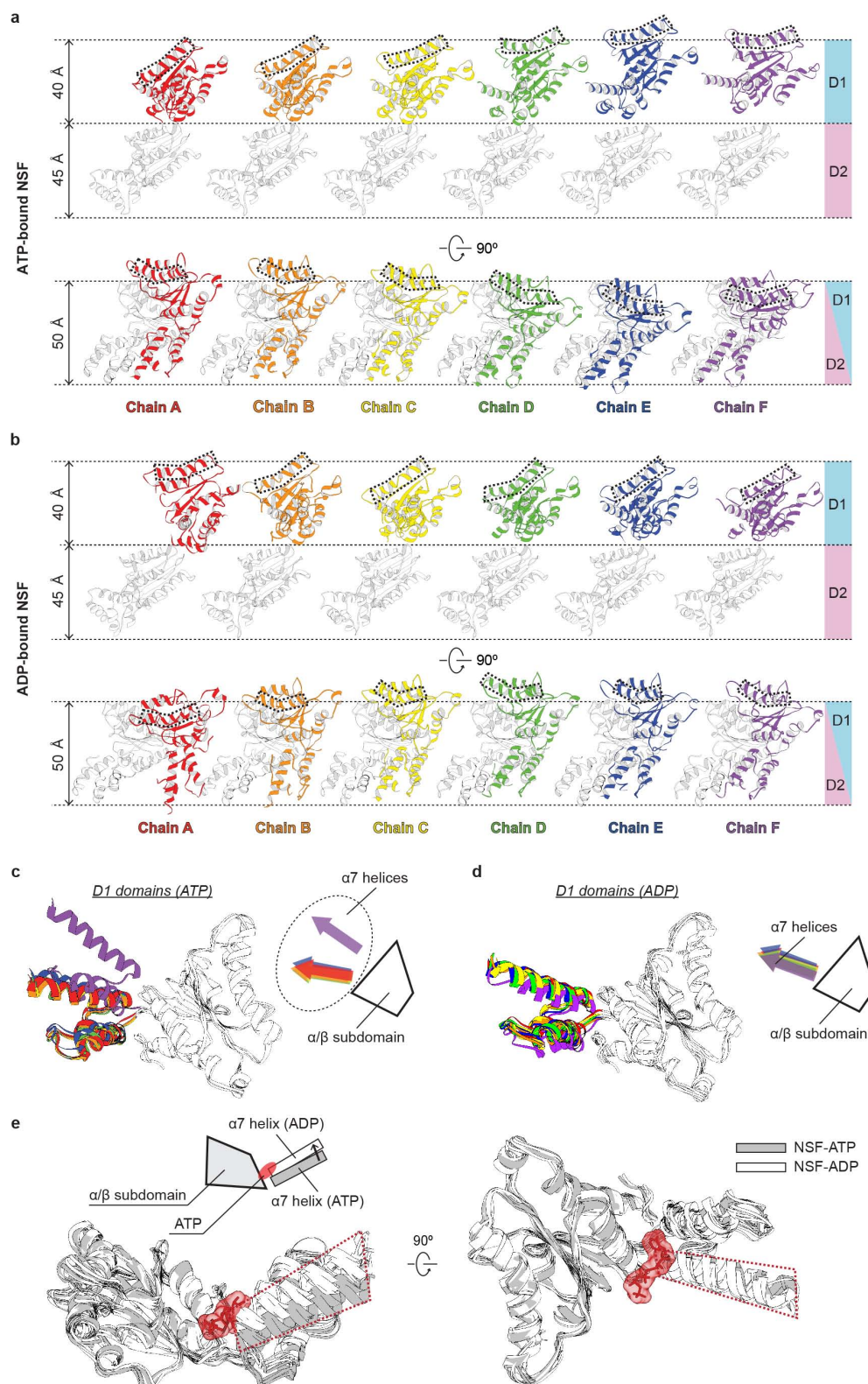
Extended Data Figure 6 | Comparison of AAA+ ATPase domains from ATP- and ADP-bound NSF structures. **a**, Unrolling of the ATPase domains of ATP-bound NSF. Two orthogonal views are shown. Individual chains are aligned based on the D2 domains (white) to show the split-washer arrangement of the D1 domains. **b**, Unrolling of the ATPase domains of ADP-bound NSF. Individual chains are aligned as in panel **a**. Dotted boxes in panels **a** and **b** highlight the α2 helices of the D1 domains in order to help with visualization of the relative positions. The six protomer chains are rainbow coloured as in Fig. 2. **c**, Superposition of the six D1 domains of ATP-bound NSF based on the α/β subdomains (white). The relative positions of α7 helices from the α subdomains are illustrated in the inset. **d**, Corresponding superposition of the ADP-bound NSF D1 domains. **e**, Superposition of the five D1 domains (without chain F) of ATP-bound NSF (grey), and six D1 domains of ADP-bound NSF (white) based on the α/β subdomains. The α7 helices from the α subdomains are highlighted by red dotted boxes. The relative translation of the α7 helices between the ATP-bound state and the ADP-bound state is shown in the inset.

ATP-bound   *ADP-bound*

a

1372.5 Å²   1394.5 Å²   1366.9 Å²   1325.3 Å²

1316.2 Å²   1313.0 Å²

1362.7 Å²   1205.6 Å²

1442.1 Å²   1336.8 Å²   1477.0 Å²   184.7 Å²

Φmax = 124 Å   *Φmax* = 125 Å   *D2*

b

1294.2 Å²   684.5 Å²

1240.7 Å²   747.8 Å²   678.0 Å²

1178.0 Å²

1113.8 Å²   692.0 Å²

447.1 Å²

663.9 Å²   943.8 Å²

Φmax = 122 Å   *Φmax* = 140 Å   *D1*

c

90°   90°

180°   180°

N

d

NSF-ATP
20S-AMPPNP

180°

D1

D2

e

A

B

Interface area change over NSF-ATP (%)

C

F

D   E

-0.08%
+3.3%
+2.0%
+12%
+0.01%
+0.27%

D2

f

A   α7 (NSF-ATP)

B   F

α7 (20S-AMPPNP)

C   E

D

-16%
-0.06%
+3.5%
-2.9%
+32%
+15%

D1

Extended Data Figure 7 | Comparison of ATP- and ADP-bound NSF structures, and ATPase domains of ATP-bound NSF and 20S supercomplex. a–c, Surface representations of the D2, D1 and N domains of ATP- and ADP-bound NSF (looking down from the N-terminal side of the NSF hexamer). The maximum diameters of the D2 and D1 rings, and the interface areas (calculated by PISA[66]) between ATPase domains are indicated. Each protomer chain is coloured as in Fig. 2. The D1 ring is also shown in panel **c** and coloured white to help with visualization. **d–f,** The ATPase domains of the structure of the 20S supercomplex (state I) were superposed on the ATP-bound NSF using the D1 ring as the reference for the fit. Six protomer chains from ATP-bound NSF are rainbow coloured as in Fig. 2. The ATPase domains of the 20S supercomplex are colored in white and grey. Note that the density of chain F in the reconstruction of ATP-bound NSF alone is poorly resolved (Fig. 1b), whereas in the 20S reconstruction it is well defined, although the overall resolution of the 20S reconstruction is lower. **d,** Side views. **e,** Top view of the D2 rings. Each individual D2 domain is labelled. Percentages of interface area change (from NSF to 20S) between the D2 domains are provided in the figure. The interface areas between the D2 domains are similar in the NSF and 20S structures, except for a significant increase (12%) between chains D and E for 20S compared to NSF alone. **f,** Top view of the D1 rings. Each D1 domain is labelled, with the split between chains A and F indicated by a black arrow. The translation of the α7 helix in α subdomain of chain A is illustrated in the inset. Percentages of interface area change (from NSF to 20S) between the D1 domains are shown. Three interfaces stay the same; the one between chains A and B decreases, whereas those between chains E and F, and chains F and A increase significantly.
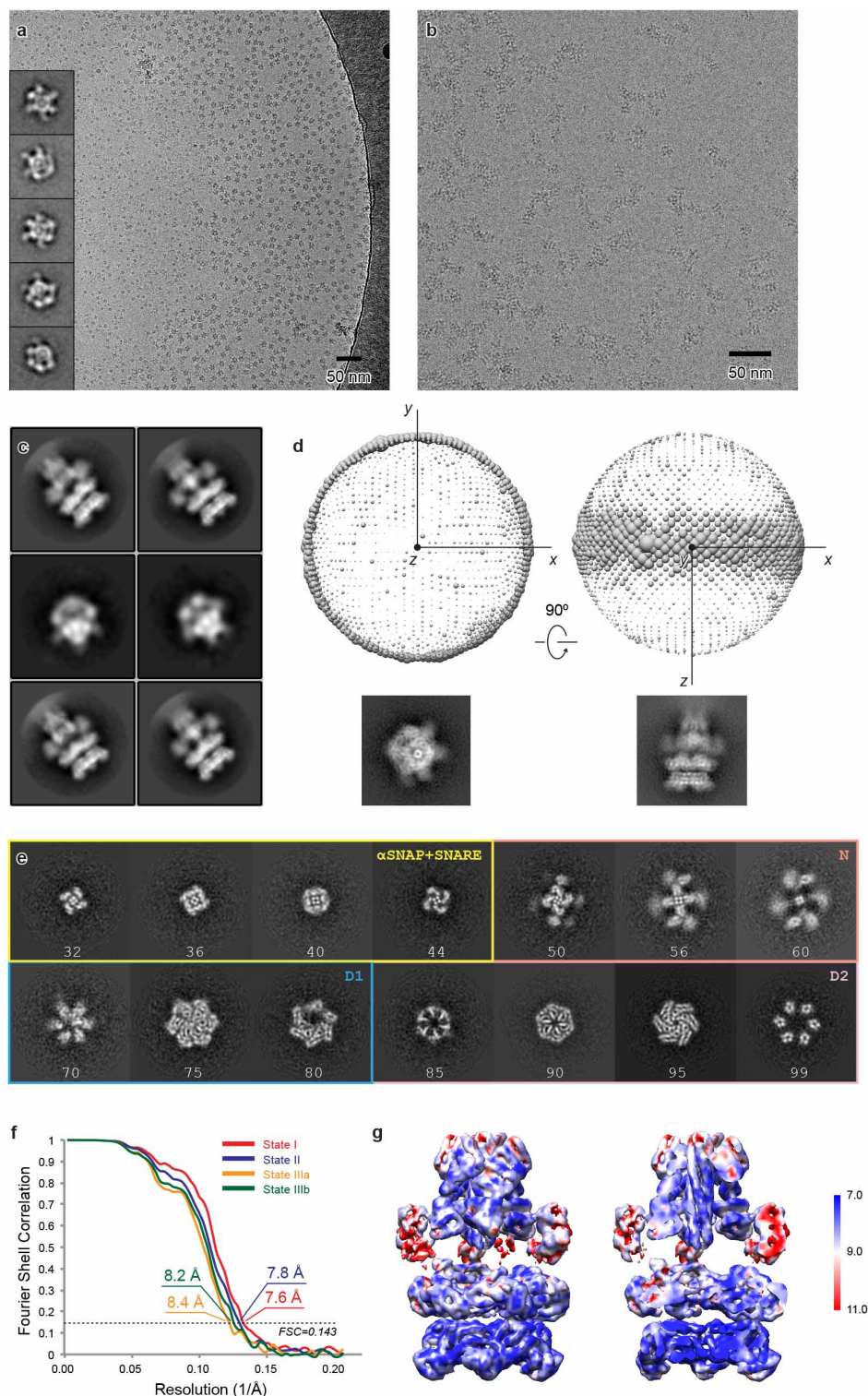
Extended Data Figure 8 | 3D reconstruction of 20S supercomplex by single-particle cryo-EM. a, A representative electron micrograph (out of 610 micrographs) of the 20S supercomplex in its original purification buffer recorded using the TF20 microscope and the TVIPS TemCam-F816 CMOS camera. The inset shows selected 2D class averages (5 out of 50). **b**, A representative electron micrograph (out of 2,459 micrographs) of 20S supercomplex in the buffer containing additional 0.05% Nonidet P-40 recorded using the TF30 Polara microscope and the Gatan K2 Summit detector. **c–g**, Results from this imaging condition. **c**, Selected 2D class averages (6 out of 50). **d**, Plots of angular distribution of particle projections. The radius of the sphere at each projection direction is proportional to the number of particle images assigned to it. Two alternative views are shown, with either the *z* axis or the *y* axis pointing out towards the viewer. Two corresponding re-projection images of the final density map are shown under the plots. **e**, Selected slice views of the final reconstruction. Slice numbers are indicated. Slices from different layers are framed in different colours: SNAREs and αSNAPs: yellow, N domains: pink, D1 ring: blue, and D2 ring: purple. **f**, FSC curves for the 3D density maps of the four states after RELION post-processing. The estimated resolution ranges from 7.6 Å to 8.4 Å as estimated by the gold-standard refinement criterion. **g**, 3D densi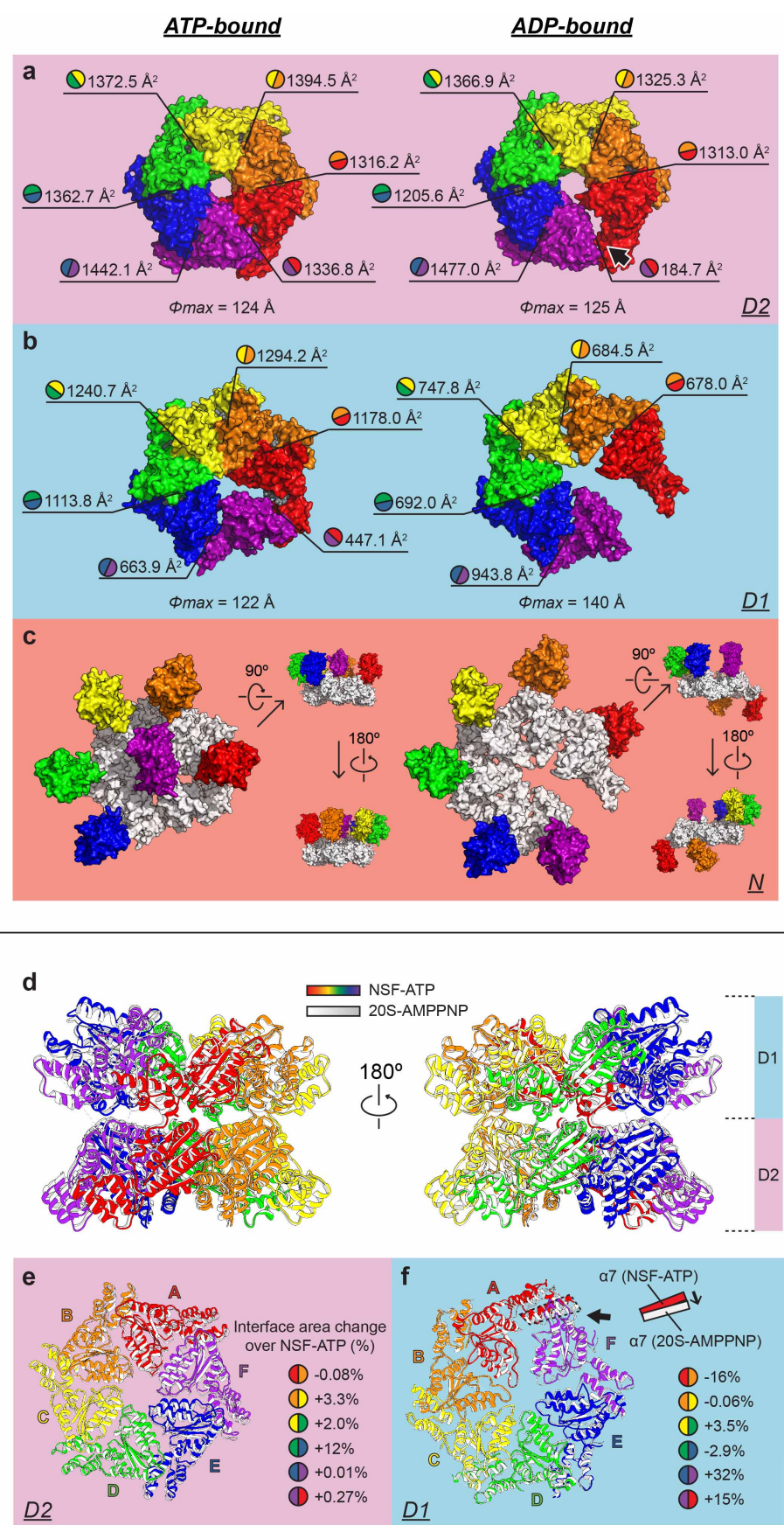ty map coloured using local resolution estimated by ResMap. The right panel shows a cut-through view of the interior of the map. **c–e** and **g** are results from a subclass representing state I.
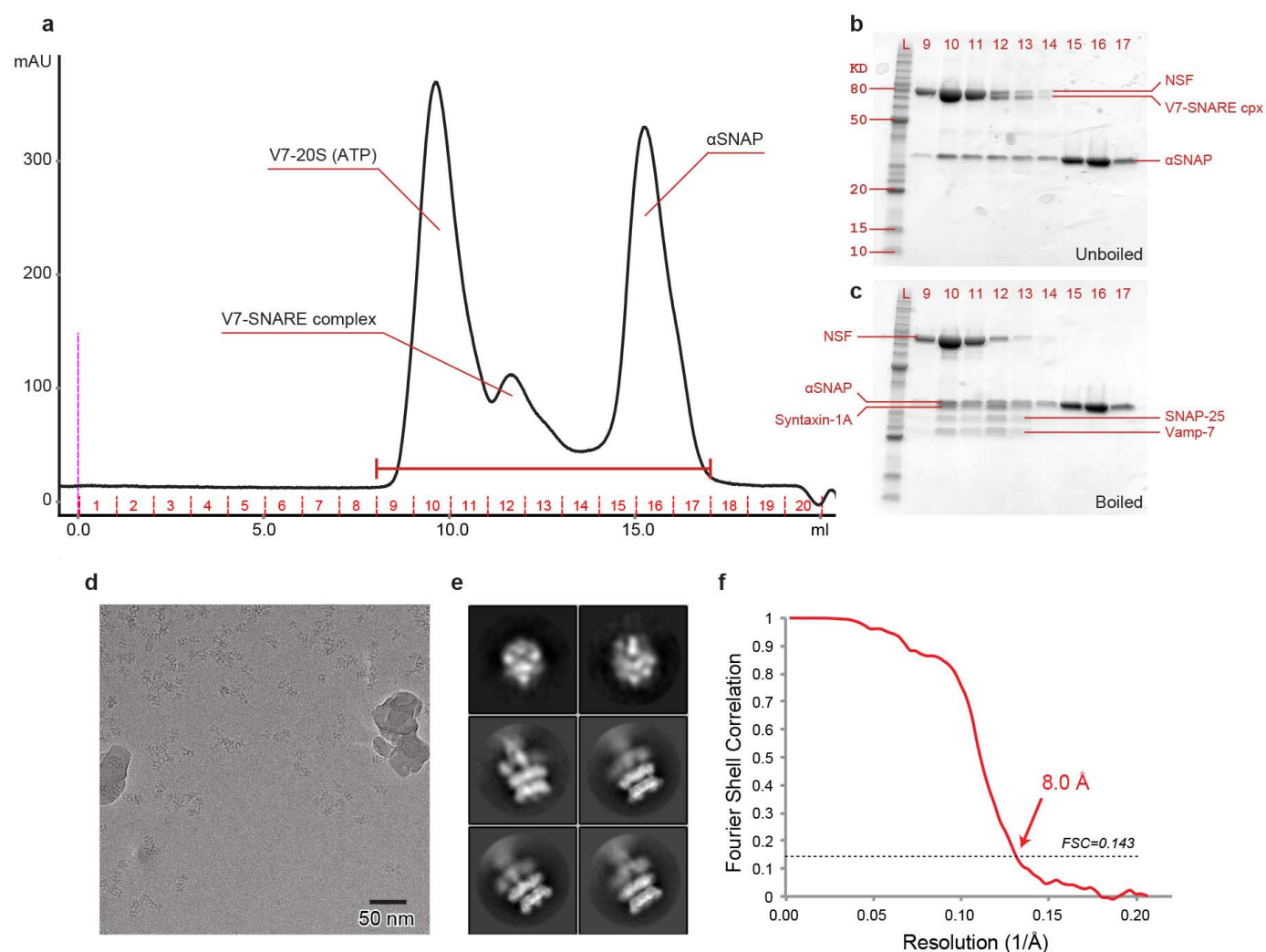
Extended Data Figure 9 | Purification and 3D reconstruction of V7-20S supercomplex. a, Size-exclusion chromatogram of the V7-20S supercomplex. Major peaks are labelled. Only fraction 10 was concentrated and used for single-particle cryo-EM. **b**, SDS–PAGE gel of fractions collected in panel **a**. The samples were not boiled. **c**, SDS–PAGE gel of the same fractions as in panel **b**. The samples were boiled. **d**, A representative electron micrograph (out of 993 micrographs) of the V7-20S supercomplex. **e**, Selected 2D class averages (6 out of 50). **f**, FSC curve for the 3D density map after RELION post-processing. The estimated resolution is 8.0 Å as estimated by the gold-standard refinement criterion.
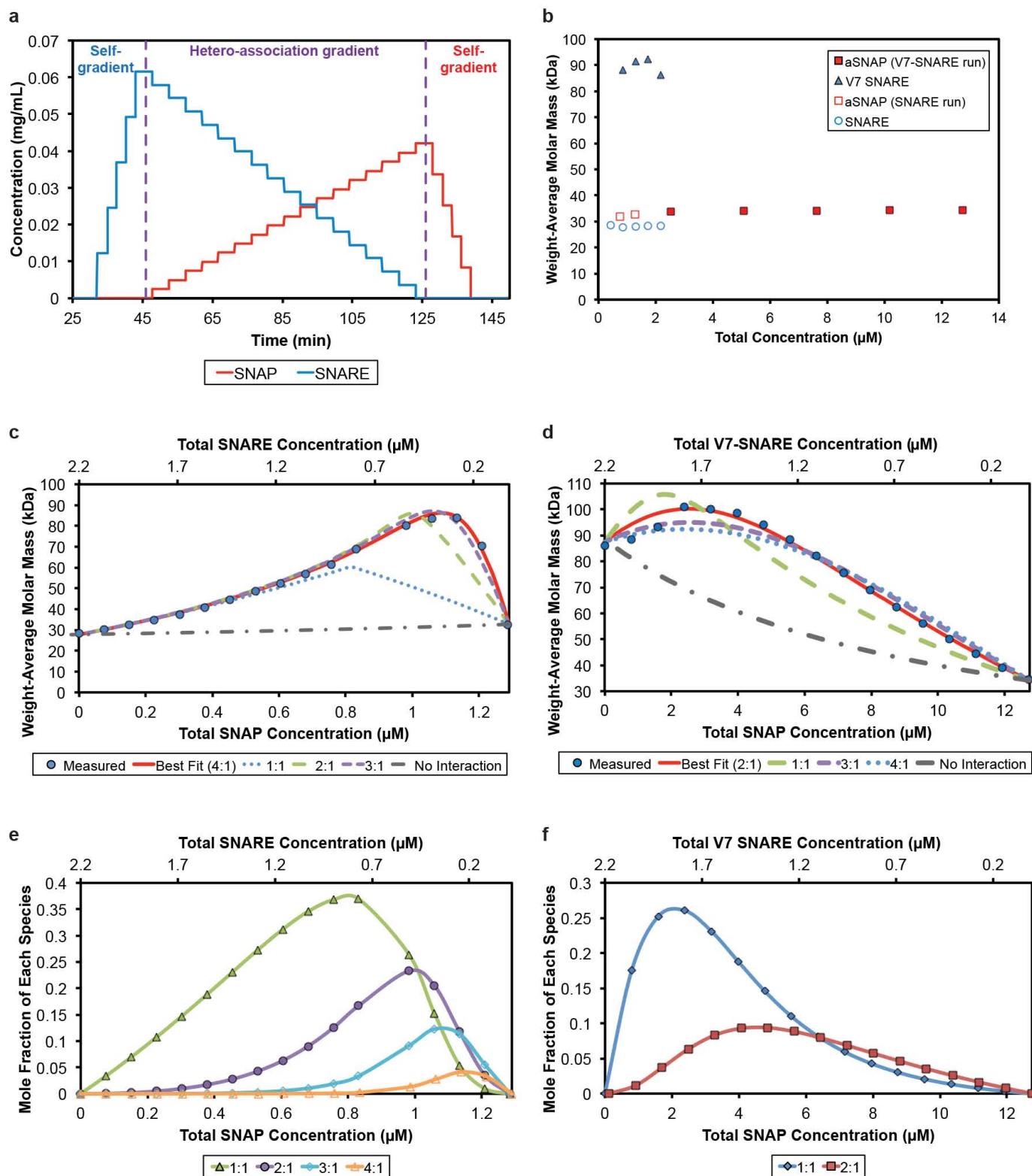
Extended Data Figure 10 | CG-MALS characterization of αSNAP–SNARE subcomplex. a, Concentration gradient setup for the experiment that measures the binding between αSNAP and truncated neuronal SNARE complex. **b,** Measured molar mass for different components. Note that there were two independent runs for αSNAP over the specified concentration ranges. **c,** Measured molecular mass of αSNAP–SNARE (truncated) subcomplex converted from light scattering over the concentration gradient. The experimental data are represented by blue dots. Simulated curves with different αSNAP:SNARE (truncated) stoichiometry are shown. The best fit is 4:1.

d, Measured molecular mass of the αSNAP-V7-SNARE subcomplex calculated from light scattering over the concentration gradient. The experimental data are represented by blue dots. Simulated curves with different αSNAP:V7-SNARE stoichiometry are shown. The best fit is 2:1. **e,** Calculated mole fractions of different αSNAP–SNARE (truncated) species over the concentration gradient based on 4:1 stoichiometry. **f,** Calculated mole fractions of different αSNAP–V7-SNARE species over the concentration gradient based on 2:1 stoichiometry.

Extended Data Table 1 | 3D reconstructions of NSF and 20S supercomplex by single particle cryo-EM

	NSF (ATP)	NSF (ADP)		20S			V7-20S
Electron microscope	TF30 Polara	TF30 Polara		TF30 Polara			TF30 Polara
Accelerating voltage (kV)	300	300		300			300
Defocus range (μm)	-1.8 − -2.8	-1.8 − -2.8		-1.8 − -2.8			-1.8 − -2.8
Electron dose ($e^-/\text{Å}^2$)	44 (26.4*)	44		44			44
Pixel size (Å)	1.2156	2.4312		2.4312			2.4312
Particles processed	89,289	30,848		116,082			65,126
			state I	**state II**	**state III$_a$**	**state III$_b$**	
Particles refined	50,781	12,830	29,717	21,489	15,249	14,991	32,100
Resolution of unmasked map (Å)	6.4	9.2	8.6	8.9	9.4	9.2	9.2
Resolution of masked map (Å)†	4.2	7.6	7.6	7.8	8.4	8.2	8.0
Map sharpening B-factor (Å^2)‡	-129	-479	-428	-601	-612	-734	-395

* The accumulated dose of the first 18 frames.
† Resolution of the masked map is estimated from masking-effect-corrected FSC curves.
‡ B-factors automatically determined by RELION.

Extended Data Table 2 | Statistics of model refinement

	NSF (ATP)	NSF (ADP)	20S supercomplex			
Model composition						
Total atoms	21,712	21,407	41,435			
Peptide chains	6	6	14			
Protein residues	2,906	2,887	5,431			
Refinement						
Unit cell (P1)						
$a = b = c$ (Å)	311.2	311.2	311.2			
$\alpha = \beta = \gamma$ (°)	90	90	90			
			state I	**state II**	**state III$_a$**	**state III$_b$**
Resolution (Å)	4.2	7.6	7.6	7.8	8.4	8.2
R.m.s. deviations						
Bond lengths (Å)	0.008	0.011	0.009	0.009	0.009	0.009
Bond angels (°)	1.762	1.976	1.596	1.567	1.585	1.570
Ramachandran plot						
Favored (%)	93.0	92.8	89.6	88.9	89.3	89.2
Outliers (%)	2.0	2.2	2.3	2.5	2.3	2.3
MolProbity						
Overall score	2.62	2.54	2.54	2.58	2.59	2.57
Rotamer outliers (%)	1.8	0.8	0.9	1.0	1.1	1.1

ARTICLE

Transport domain unlocking sets the uptake rate of an aspartate transporter

Nurunisa Akyuz[1], Elka R. Georgieva[2,3], Zhou Zhou[1], Sebastian Stolzenberg[1], Michel A. Cuendet[1,4], George Khelashvili[1], Roger B. Altman[1], Daniel S. Terry[1], Jack H. Freed[2,3], Harel Weinstein[1,5], Olga Boudker[1] & Scott C. Blanchard[1,6]

Glutamate transporters terminate neurotransmission by clearing synaptically released glutamate from the extracellular space, allowing repeated rounds of signalling and preventing glutamate-mediated excitotoxicity. Crystallographic studies of a glutamate transporter homologue from the archaeon *Pyrococcus horikoshii*, Glt_{Ph}, showed that distinct transport domains translocate substrates into the cytoplasm by moving across the membrane within a central trimerization scaffold. Here we report direct observations of these 'elevator-like' transport domain motions in the context of reconstituted proteoliposomes and physiological ion gradients using single-molecule fluorescence resonance energy transfer (smFRET) imaging. We show that Glt_{Ph} bearing two mutations introduced to impart characteristics of the human transporter exhibits markedly increased transport domain dynamics, which parallels an increased rate of substrate transport, thereby establishing a direct temporal relationship between transport domain motion and substrate uptake. Crystallographic and computational investigations corroborated these findings by revealing that the 'humanizing' mutations favour structurally 'unlocked' intermediate states in the transport cycle exhibiting increased solvent occupancy at the interface between the transport domain and the trimeric scaffold.

Glutamate transporters, also termed excitatory amino acid transporters (EAATs), maintain glutamate concentration gradients across the cell membrane by coupling neurotransmitter uptake to symport of three sodium (Na^+) ions and a proton and counter-transport of a potassium ion[1]. Structural information on the EAAT family principally stems from investigations of Glt_{Ph}[2–6], an aspartate and Na^+ symporter[7,8] from *Pyrococcus horikoshii.* Crystal structures of Glt_{Ph} revealed that the homotrimeric protein is composed of a rigid, central trimerization scaffold that supports three peripheral transport domains containing the substrate binding sites (Fig. 1a). Comparison of Glt_{Ph} structures captured in distinct conformations suggests that within the trimerization scaffold, individual transport domains undergo relocations approximately 15 Å perpendicular to the membrane, providing substrate and ions alternating access to the extracellular (outward) and intracellular (inward) solutions (Extended Data Fig. 1a)[5].

Single-molecule imaging of Glt_{Ph} provided direct evidence for bidirectional elevator-like transport domain motions[9,10]. Consonant with findings obtained using double electron-electron spin resonance (DEER) spectroscopy[11,12], these measurements also showed that individual Glt_{Ph} transport domains transition spontaneously between outward- and inward-facing conformations both when free of cargo (apo) and when bound to substrates. These transport domain motions exhibited heterogeneous dynamic behaviours, alternating between periods of rapid transitions and periods of quiescence, where the protein rests in either outward- or inward-facing states[9]. In contrast to observations in structurally unrelated neurotransmitter sodium symporters[13], substrate binding decreased transport domain dynamics in Glt_{Ph} by favouring the quiescent periods such that the frequency of domain motions converged to the substrate uptake rate[7,9].

These findings led to the hypothesis that Glt_{Ph} configurations observed in crystal structures[2,4], showing tight lock-and-key interactions between transport and trimerization domains, represent quiescent locked states with high substrate affinity, whereas the short-lived states sampled during dynamic periods are structurally distinct and likely have intrinsically lower substrate affinity (Extended Data Fig. 1b). This model posits that transport domain motions require a rate-limiting, structural unlocking process that changes the interface between the transport and trimerization domains, probably enabling solvent penetration into that interface[5,9,14].

To assess the relationship between Glt_{Ph} function, dynamics and structure, we employed smFRET imaging in the context of reconstituted proteoliposomes with physiological ion gradients. We compared wild-type Glt_{Ph} to a gain-of-function, humanized (H) mutant R276S/M395R ($H_{276,395}$-Glt_{Ph}), which through unknown mechanisms exhibits a faster rate of substrate uptake[15]. The smFRET experiments revealed that the mutations destabilized quiescent locked states. The resulting increase in dynamics paralleled a decreased affinity for substrate and an increased transport rate. Crystallographic analyses supported this observation, showing that the transport domains of $H_{276,395}$-Glt_{Ph} can adopt inward-facing conformations in which the transport domain-trimerization scaffold interface is strikingly more open than previously observed. Computational modelling further suggested that increased solvation by lipid or detergent hydrophobic tails in this interface probably facilitates the formation of such conformations. These observations provide a structural rationale for functional distinctions between Glt_{Ph} and the human EAATs, and establish a kinetic framework for understanding how regulation can be achieved.

Experimental design

Glt_{Ph} is a structural homologue of EAATs (~35% sequence identity)[2] that preferentially transports aspartate over glutamate, with higher substrate binding affinity and slower uptake rate[3,7]. It has been suggested[15]

Figure 1 | Transport rates and 'elevator-like' domain dynamics are correlated. a, Surface representation of the outward-facing Glt_Ph showing the transport (blue) and scaffold (beige) domains. In one protomer, HP1 (yellow) and TM8 (dark blue) are emphasized as cartoons. In the enlarged substrate-binding site (right), mutated residues and aspartate are shown as sticks and ions are shown as spheres. **b**, Sequence alignment for HP1 and TM8 with mutation sites highlighted in pink. **c**, Aspartate uptake by unlabelled (black)

and labelled (red) transporters. Substrate uptake data are shown as averages of at least three experiments with error bars representing standard deviations. **d**, Proteoliposome attachment strategy. **e**, smFRET trajectories recorded for the wild type (WT) and mutant in proteoliposomes under transport conditions. **f**, Transition density plots corresponding to **e**. The average transition frequencies and the number of total transitions (N_t) are shown. Colour scale is from tan (lowest) to red (highest frequency).

that these distinctions may stem, in part, from the differential location of a conserved arginine residue[16] that is proximal both to the substrate-binding site and the transport domain–trimerization scaffold interface. Although the location of this arginine can differ in the primary sequences of glutamate transporter homologues, its position is conserved in most family members (Fig. 1a, b, and Extended Data Fig. 1c). In human EAAT1, moving this arginine from transmembrane segment (TM) 8 to helical hairpin (HP) 1 (where it is located in Glt_Ph), strikingly increases substrate affinity and decreases uptake rate[15]. Reciprocal mutagenesis of Glt_Ph, whereby the arginine is moved from HP1 to TM8 (R276S/M395R), reduces aspartate affinity and increases the transport rate[15]. We took advantage of this gain-of-function mutant to probe correlations between uptake rate and transport domain dynamics.

We performed a comparative analysis of wild-type and $H_{276,395}$-Glt_Ph proteins using smFRET, in which elevator-like transport domain motions in individual Glt_Ph trimers, bearing one donor and one acceptor fluorophore were revealed as time-dependent changes in FRET efficiency[9] (Extended Data Fig. 2a, b). To investigate such motions in the context of proteoliposomes, we labelled Glt_Ph proteins with intramolecularly stabilized derivatives of the cyanine fluorophores Cy3 and Cy5 that exhibit intrinsically enhanced brightness and photostability[17,18] (Extended Data Fig. 2c), eliminating the need for fluorophore protective agents that disturb lipid bilayer properties[19]. The labelled proteins were reconstituted into liposomes in the absence of substrates for smFRET and bulk substrate uptake measurements. Bulk radioactive aspartate uptake experiments confirmed that both the labelled wild-type and $H_{276,395}$-Glt_Ph mutant transported substrate with rates similar to those of the unlabelled proteins, the mutant being about fourfold faster than the wild type (Fig. 1c).

For smFRET measurements, reconstitution procedures were established to yield maximally one Glt_Ph trimer per vesicle. Proteoliposomes were immobilized via biotinylated, fluorescently labelled Glt_Ph within passivated quartz microfluidics chambers activated with a biotin–streptavidin bridge (Fig. 1d). Using this strategy, only those proteoliposomes containing Glt_Ph oriented with the extracellular side facing the vesicle exterior were immobilized and imaged[20]. Imaging experiments were initiated in the absence of substrates under isoelectric conditions and chemical gradients were established by rapidly exchanging the proteoliposomes into a buffer containing Na^+ ions and aspartate. Additionally, we examined the behaviours of the labelled proteins in detergent micelles that afford higher signal-to-noise ratios and increased sample size.

Transport rate and dynamics are correlated

In both the absence and presence of gradients, wild-type Glt_Ph in proteoliposomes showed spontaneous transitions between low-, intermediate- and high-FRET efficiency states centred at ~0.4, ~0.6 and ~0.9, respectively (Fig. 1e, f and Extended Data Fig. 3). In detergent solutions, these FRET states were assigned to specific Glt_Ph configurations: the low-FRET state reflects symmetric outward-facing and asymmetrically outward- and inward-facing configurations; intermediate- and high-FRET states reflect, respectively, asymmetrically inward- and outward-facing and simultaneously inward-facing protomers (Extended Data Fig. 2)[9]. In line with previous investigations[9,11,12], population FRET data from hundreds of individual proteins in the absence of gradients show that the transporter occupies the outward-facing, low-FRET state about half of the time in both detergent (46%) and lipid vesicles (54%) (Fig. 2a, Extended Data Fig. 3b and Extended Data Table 1a, b).

Figure 2 | Ligand-dependent state distributions in detergent. a, c, In each panel, time-dependent FRET efficiency population contour plots (left) and cumulative population histograms (right) are shown for the wild type (**a**) and mutant (**c**). Experimental conditions are indicated abov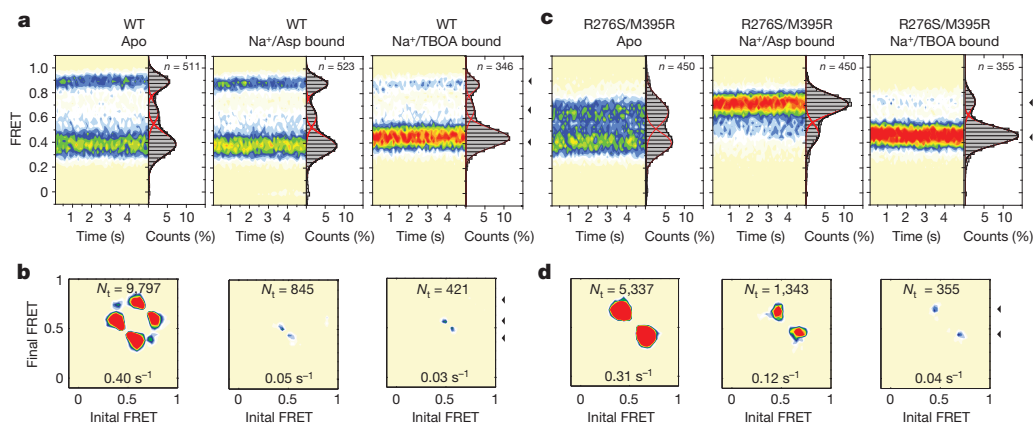e the panels. Contour plots are colour-coded from tan (lowest) to red (highest population); colour scale from 0–12%. Histograms display the time-averaged state distributions. Solid black lines are fits to sums of individual Gaussian functions (red lines). *n* is the number of molecules analysed. **b, d,** Corresponding transition density plots (as in Fig. 1).

Transitions between low- and higher-FRET states reflect elevator-like movements of the individual transport domains between outward- and inward-facing configurations, respectively[9]. In proteoliposomes, such transitions occurred at a rate of ~0.2 s⁻¹, roughly twofold less frequently than in detergent (Fig. 2a and Extended Data Fig. 3c). Paralleling the effects of substrate binding to wild-type Glt$_{Ph}$ in detergent (Fig. 2b), a modest population shift towards the outward-facing, low-FRET state occurred under active transport conditions achieved by addition of Na⁺ and aspartate (Extended Data Fig. 3b and Extended Data Table 1a, b). Na⁺ and aspartate also reduced transport domain dynamics by tenfold to ~0.02 s⁻¹ (Fig. 1f and Extended Data Fig. 3c). Thus, in the presence of chemical gradients, the frequency of transitions from outward- to inward-facing state (~0.01 s⁻¹) mirrored the rate of radioactive substrate uptake (~0.007 s⁻¹) (Fig. 1c).

Notably, the H$_{276,395}$-Glt$_{Ph}$ mutant only exhibited transitions between low- (0.4) and a single, higher- (0.65) FRET state in both proteoliposomes and detergent (Figs 1f and 2c and Extended Data Fig. 3). Similar to the wild-type protein in the absence of chemical gradients, the low-FRET state was occupied 60% of the time in detergent micelles and 40% in proteoliposomes (Extended Data Table 1a, b). The observed FRET transition frequency for H$_{276,395}$-Glt$_{Ph}$ was also two times slower in proteoliposomes (~0.13 s⁻¹) compared to detergent (~0.22 s⁻¹) (Fig. 2d and Extended Data Fig. 3c).

However, in stark contrast to the wild-type protein, the transition frequency in H$_{276,395}$-Glt$_{Ph}$ decreased by less than twofold to ~0.1 s⁻¹ when transport-supporting chemical gradients were established (Fig. 1e, f and Extended Data Fig. 3c). Here again, the frequency of transitions from the outward- to inward-facing FRET state (~0.05 s⁻¹) converged to the measured rate of substrate uptake (~0.03 s⁻¹) (Fig. 1c). The quantitative correspondence observed between the rates of smFRET transitions and uptake for the wild-type and H$_{276,395}$-Glt$_{Ph}$ mutant proteins provides compelling evidence that elevator-like motions of transport domains mediate solute uptake and are critical steps of the transport cycle[9]. This finding was independent of the proteoliposome immobilization strategy used and valinomycin-mediated electrical potentials (Extended Data Fig. 4a–d)[10].

H$_{276,395}$-Glt$_{Ph}$ visits distinct inward-facing states

In contrast to wild-type Glt$_{Ph}$, which samples intermediate- (0.6) and high- (0.9) FRET states, H$_{276,395}$-Glt$_{Ph}$ samples only a single higher-FRET (0.65) configuration (Extended Data Fig. 3). No excursions into the 0.9 FRET state were observed even when data were collected at sixfold higher time resolution (15 ms) (Extended Data Fig. 4e, f). The absence of the 0.9 FRET state would be expected if only one protomer within the H$_{276,395}$-Glt$_{Ph}$ trimer transitioned into inward-facing configuration at a time, while the formation of symmetric inward-facing states were disallowed. This model is, however, inconsistent with data

showing that individual transport domains function independently[5,14,21,22]. An alternative hypothesis is that the inward/outward and inward/inward configurations in H$_{276,395}$-Glt$_{Ph}$ exhibit altered, overlapping FRET values. If this model is correct, then the gain-of-function mutations in H$_{276,395}$-Glt$_{Ph}$ have altered the nature of the elevator-like transport domain motions and the structure of the inward-facing state.

The energy landscape of H$_{276,395}$-Glt$_{Ph}$ is altered

Na⁺ and aspartate significantly stabilized the higher-FRET state of H$_{276,395}$-Glt$_{Ph}$ in detergent micelles (Fig. 2c and Extended Data Table 1a, b). In detergent, Na⁺ and aspartate have access to both the extracellular and cytoplasmic sides of the protein. Assuming that a binding equilibrium is established in each conformation, these observations suggest that substrates bind more tightly to the inward-facing H$_{276,395}$-Glt$_{Ph}$ conformation. Such a response was not observed for the wild-type Glt$_{Ph}$, where substrate affinities of the inward- and outward-facing conformations are nearly the same[23] and ligands stabilize the latter only slightly[9,11,12] (Fig. 2a and Extended Data Table 1). Notably, the transporter blocker DL-threo-β-benzyloxyaspartate (TBOA)[24] stabilized the outward-facing low-FRET states of both wild-type and H$_{276,395}$-Glt$_{Ph}$ (Fig. 2 and Extended Data Fig. 5a, b). As above, this suggests that TBOA preferentially binds to the outward-facing state of both isoforms. Results consistent with these findings were obtained from ensemble DEER measurements using the protein spin-labelled on the same residue (Extended Data Fig. 5c).

Interestingly, the addition of Na⁺ and aspartate to H$_{276,395}$-Glt$_{Ph}$ proteoliposomes led to an increase in the outward-facing, low-FRET
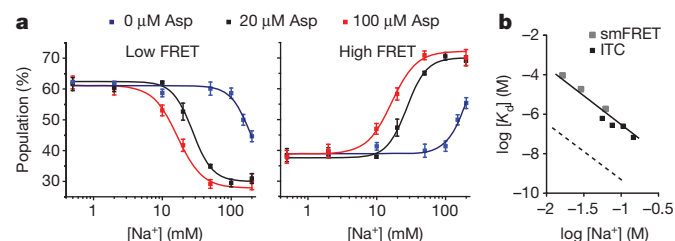
Figure 3 | Coupled Na⁺ and aspartate binding to H$_{276,395}$-Glt$_{Ph}$.
a, Populations of low- (left) and higher- (right) FRET states determined for H$_{276,395}$-Glt$_{Ph}$ in detergent micelles as functions of Na⁺ concentration in the presence of 0 (blue), 20 (black) and 100 μM (red) aspartate. Solid lines are fits to Hill equation with K_d = 200, 30 and 15 mM, respectively, and *n* value of 3. The data points shown are averages and standard errors from three independent biological replicates. **b,** Logarithmic plots of aspartate K_d values as functions of Na⁺ concentrations. Data are from isothermal titration calorimetry (ITC) (black) and smFRET (grey). The solid line through the data are a linear fit with slope 3.2. The extent of coupling between Na⁺ and aspartate binding is similar to wild type (dashed line)[23].
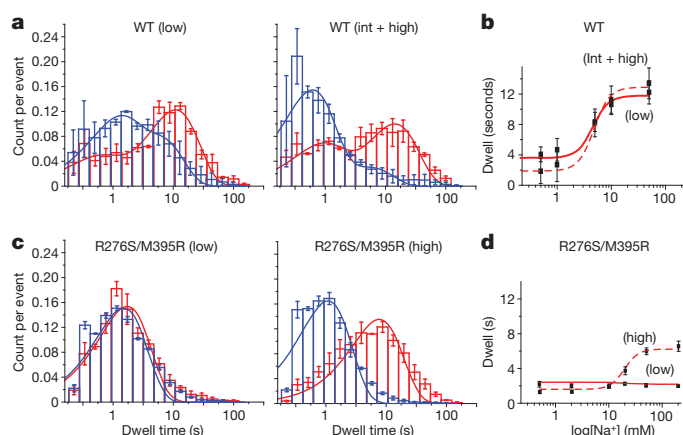
Figure 4 | Unimodal dynamic behaviour of $H_{276,395}$-Glt$_{Ph}$. a–d, Dwell time distributions (**a, c**) and average dwell times (**b, d**) for the low- (low, solid lines) and intermediate- and high-FRET states (int + high, dashed lines) obtained for the wild type (**a, b**) and $H_{276,395}$-Glt$_{Ph}$ (**c, d**) in detergent. The distributions for apo (blue) and Na$^+$/aspartate-bound proteins (red) were fitted to a probability density function. The fitted time constants are in Extended Data Table 1c. Average dwells are plotted as functions of Na$^+$ concentration in the presence of 10 and 100 μM aspartate for wild type and $H_{276,395}$-Glt$_{Ph}$, respectively. Solid lines are fits to Hill equation with K_d = 15 mM and n = 3.2 for wild type and K_d = 19 mM and n = 3.2 for $H_{276,395}$-Glt$_{Ph}$. The data points shown are averages and standard errors from three independent biological replicates.

population (Extended Data Fig. 3). This liposome-specific response to substrate addition provides supporting evidence for bilayer integrity. It also reveals that elevator-like transport domain movements—as opposed to substrate release—are rate-limiting in the $H_{276,395}$-Glt$_{Ph}$ transport cycle. If substrate release were slow compared to the domain movements, the state distributions during uptake would mirror those observed in detergent, that is, show preference for the inward-facing higher-FRET state.

The effect of substrate on the distribution of FRET states observed for both isoforms was concentration-dependent in detergent micelles.

Notably, $H_{276,395}$-Glt$_{Ph}$ exhibited an ∼1,000-fold increase in apparent substrate dissociation constant (K_d) compared to the wild-type protein (Fig. 3 and Extended Data Fig. 6a). This finding was corroborated by bulk measurements (Extended Data Fig. 6b, c). Hence, the $H_{276,395}$-Glt$_{Ph}$ mutations affect both transport domain dynamics and substrate affinity, even though neither of the mutated residues coordinates aspartate directly in the existing crystal structures. These observations support the hypothesis that substrate binding and transport domain dynamics are physically coupled.

Locked states are destabilized in $H_{276,395}$–Glt$_{Ph}$

The coexistence of quiescent and dynamic periods evidenced both in the absence and presence of ligands is a hallmark kinetic feature of wild-type Glt$_{Ph}$[9]. Binding of Na$^+$ and aspartate increases the prevalence of quiescent periods and thus the average FRET state lifetimes (Fig. 4a, b). Strikingly, no evidence was found for quiescent periods in the $H_{276,395}$-Glt$_{Ph}$ mutant (Fig. 4c) and rapid transport domain dynamics persisted even in the presence of saturating substrate concentrations (Fig. 2d and Extended Data Table 1c). These dynamic processes were efficiently blocked by TBOA (Fig. 2b, d) consistent with their putative role in substrate transport. In $H_{276,395}$-Glt$_{Ph}$, substrate binding increased the lifetime of the high-FRET state (∼sevenfold), with no detectable impact on the low-FRET state lifetime (Fig. 4d). In both the absence and presence of ligands, the low- and higher-FRET state lifetimes were unimodal (Fig. 4c, d and Extended Data Table 1c). These findings suggest that in $H_{276,395}$-Glt$_{Ph}$, the isomerization steps leading to locked configurations of the wild-type protein are strikingly altered or inaccessible under the conditions examined, although an allosteric coupling between substrate binding and stabilization of the domain interface still exists.

Structure of the inward-facing $H_{276,395}$–Glt$_{Ph}$

To probe the underpinnings of the altered properties of $H_{276,395}$-Glt$_{Ph}$, we determined a crystal structure of the protein bound to Na$^+$ ions and aspartate at a moderate resolution of ∼4.5 Å (Extended Data Fig. 7). As expected from smFRET experiments (Fig. 2c), the structural model clearly showed that all protomers in the trimer spontaneously adopted
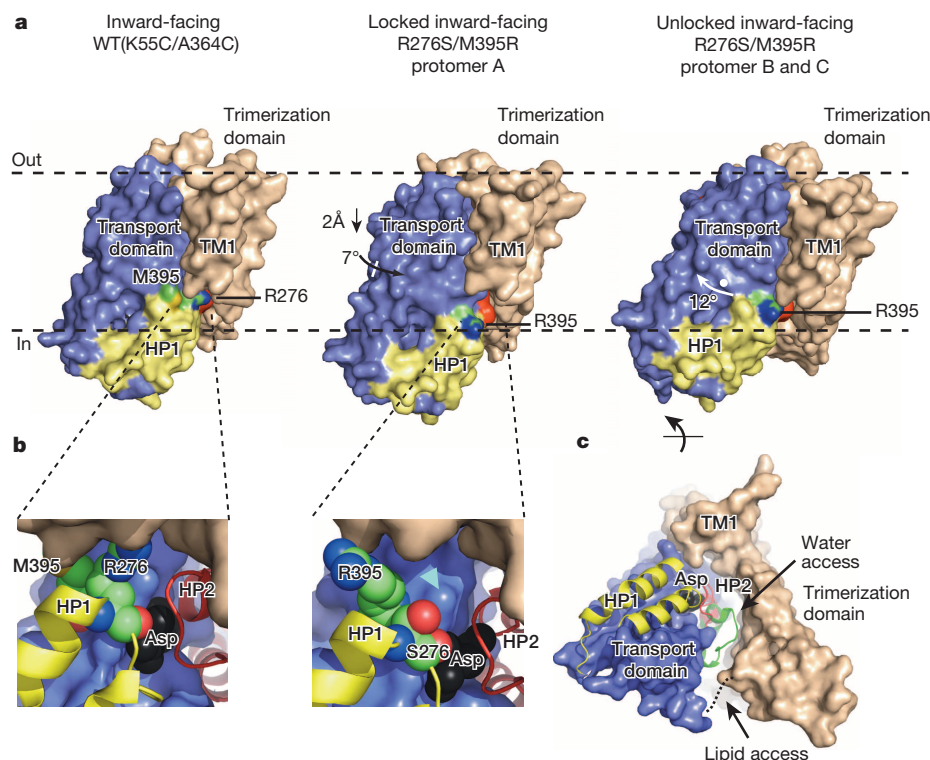
Figure 5 | Crystal structure of the $H_{276,395}$-Glt$_{Ph}$. a, Single protomers of inward-facing wild type (left), locked mutant (centre) and unlocked mutant (right) in surface representation, coloured as in Fig. 1, HP2 is red. Residues 276 and 395 are coloured by atom type. The approximate limits of the hydrocarbon layer of the membrane are shown as dashed lines. **b,** Substrate binding sites (enlarged) viewed from the cytoplasm. HP1 and HP2 are in cartoon representation; aspartate (black) and residues 276, 395 and Asp 394 (coloured by atom type) are emphasized as spheres. Arrowhead (cyan) marks the region of increased solvent accessibility. **c,** Cytoplasmic view of the unlocked protomer showing the crevice at the domain interface. Dashed line replaces TM2–TM3 loop for clarity. Arrows indicate regions of increased water and lipid accessibility. Open conformations of HP2 were modelled based on the TBOA-bound (green) structure of Glt$_{Ph}$[3].
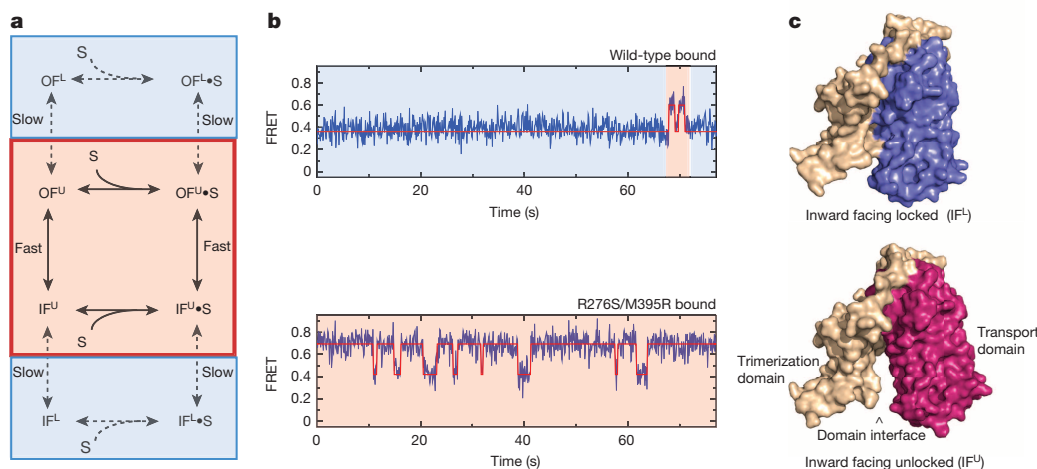
Figure 6 | Kinetic model of transport. a, Schematic representation of the transport cycle. Red and blue boxes indicate dynamic and quiescent periods, respectively. **b**, Simulated single-molecule trajectories for wild-type Glt$_{Ph}$ (top) and H$_{276,395}$-Glt$_{Ph}$ (bottom). Periods of long-lived locked states are shaded blue; periods of transitions between unlocked states are shaded red. **c**, Structures of the inward-facing wild-type locked state (top) and H$_{276,395}$-Glt$_{Ph}$ unlocked state (bottom).

inward-facing configurations. The model also revealed that the transport domain's orientations differed from those previously captured in Glt$_{Ph}$ structures, both with and without stabilizing crosslinks[4,5]. Moreover, the trimer was asymmetric, with the transport domain of protomer A occupying a position distinct from the other two.

In protomer A, the transport domain shifted further inward by 2 Å and rotates by 7° around an axis roughly perpendicular to the membrane plane with respect to the wild type (Fig. 5a). This rearrangement is accommodated by a concerted movement of helices in the scaffold domain, comprising TM1 and peripheral portions of TM2 and TM5 (Extended Data Fig. 7c), whose flexible nature was already noted[5]. This conformation resembles the inward-facing, locked state of the wild type[4,5] in the close packing observed between the transport domain and the trimerization scaffold (Extended Data Fig. 7d). Molecular dynamics simulations revealed that whereas Arg 276 in the wild type forms hydrogen bonds with Asp 394 and bulk water molecules, the corresponding Arg 395 in H$_{276,395}$-Glt$_{Ph}$ faces the hydrophobic core of the bilayer. The resulting membrane remodelling is driven by the hydrophobic matching force[25,26], and is established by interactions of the Arg 395 side chain with penetrating lipid phosphate groups and accompanying water molecules (Extended Data Fig. 8). Consequently, the penetrating polar moieties are positioned in an otherwise hydrophobic region of H$_{276,395}$-Glt$_{Ph}$, which can destabilize the inward-facing, locked conformation and increase water accessibility to the substrate-binding site and to the domain interface (Fig. 5b).

In protomers B and C, the transport domains undergo identical and more striking changes (Fig. 5a), each swinging away from the trimerization scaffold by about 12° compared to locked protomer A. Consequently, a large crevice opens between HP2 and the scaffold, reducing the interface between the transport and scaffold domains from ~1,300 Å^2 to ~900 Å^2 and allowing access to water, detergent or lipid molecules (Fig. 5c). This unusual, apparently unlocked, conformation was observed in two protomers occupying distinct crystal packing environments and therefore seems to be determined by the properties of the protein itself and not by crystal contacts. The crevice it generates is largely hydrophobic, and closes rapidly in molecular dynamics simulations when solvated only by water (Extended Data Fig. 9a–c). In contrast, the open interface between transport and trimerization domains is stable with lipids positioned in this space (Extended Data Fig. 9 d–g), suggesting that solvation by lipid or detergent molecules, is necessary. Notably, this crevice may allow HP2, whose gating role in the outward-facing state is well-established[3,6], to open when the transport domain is inward facing (Fig. 5c). If so, the substrate release might be facilitated in the unlocked conformation, a notion compatible with the markedly reduced substrate affinity of this mutant.

Discussion

Conformational transitions between outward- and inward-facing states are key events in transport cycles of secondary active transporters[27,28]. In glutamate transporters and possibly other families[29–31], such transitions involve elevator-like movements of the substrate-binding domains supported by relatively rigid scaffold domains. The frequency of such transitions in Glt$_{Ph}$ in lipid bilayers and in the presence of physiological ionic gradients parallels the turnover rate of substrate uptake. This relationship also holds in a gain-of-function mutant H$_{276,395}$-Glt$_{Ph}$ that exhibits a 1,000-fold decreased substrate affinity and a fourfold faster uptake rate. Collectively, our observations establish a direct correlation between the transport domain movements and substrate transport, and suggest an inverse relationship between substrate affinity and transport domain motions. The H$_{276,395}$-Glt$_{Ph}$ mutant is special in this regard, as other point mutations impact dynamics only and do not potentiate transport[9].

The observed dynamic signatures strongly suggest that the rate-limiting step in this process is the unlocking of the transport domain from the trimerization scaffold (Fig. 6a). Although both the wild-type and the H$_{276,395}$-Glt$_{Ph}$ proteins exhibit similar transport domain structures and translocate similarly positioned charged groups (including Arg 276 in the wild type and Arg 395 in the mutant), locked states are relatively unstable in the H$_{276,395}$-Glt$_{Ph}$ mutant, leading to overall faster dynamics and uptake.

The locked and unlocked configurations of wild-type Glt$_{Ph}$, corresponding to quiescent and dynamic periods, respectively, coexist and interconvert spontaneously, which suggests that outward- and inward-facing states of Glt$_{Ph}$—and by extension EAATs—should be viewed as structurally heterogeneous ensembles. Increased quiescent period durations in the presence of substrate further suggest that ligand binding is allosterically coupled to the formation of locked states[9]. Based on these insights, we propose a simplified kinetic framework for the transport cycle that recapitulates the most salient experimentally observed features (Fig. 6b, Extended Data Fig. 10). The specific relationship of crystallographic snapshots of Glt$_{Ph}$ and related proteins to the topological features of this framework will need to be examined carefully.

The structure of H$_{276,395}$-Glt$_{Ph}$ (Fig. 6c) captures an unlocked configuration that appears relevant to the proposed transport cycle and uniquely suitable for ligand binding and release. Although the molecular basis of how the mutations in H$_{276,395}$-Glt$_{Ph}$ affect the locked–unlocked isomerization requires further investigation, molecular dynamics simulations suggest that protein–lipid interactions are pivotal (Extended Data Fig. 9). The proposed role for the lipid hydrophobic tails in facilitating domain unlocking complements previous hypotheses that transient interface hydration facilitates transport domain translocation[5,14].

That two closely related Glt_{Ph} isoforms exhibit distinct kinetic and structural signatures foreshadows the possibility that human EAATs differ substantially from Glt_{Ph}, especially in their dynamic properties. Probing EAATs directly is therefore essential, particularly since the extent to which they might be diverted to kinetically stable, potentially off-pathway states may represent a regulatory modality.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Zerangue, N. & Kavanaugh, M. P. Flux coupling in a neuronal glutamate transporter. *Nature* **383**, 634–637 (1996).
2. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431**, 811–818 (2004).
3. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* **445**, 387–393 (2007).
4. Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880–885 (2009).
5. Verdon, G. & Boudker, O. Crystal structure of an asymmetric trimer of a bacterial glutamate transporter homolog. *Nature Struct. Mol. Biol.* **19**, 355–357 (2012).
6. Verdon, G., Oh, S., Serio, R. N. & Boudker, O. Coupled ion binding and structural transitions along the transport cycle of glutamate transporters. *ELife* **3**, e02283 (2014).
7. Ryan, R. M., Compton, E. L. & Mindell, J. A. Functional characterization of a Na+-dependent aspartate transporter from *Pyrococcus horikoshii*. *J. Biol. Chem.* **284**, 17540–17548 (2009).
8. Groeneveld, M. & Slotboom, D. J. Na$^+$:aspartate coupling stoichiometry in the glutamate transporter homologue Glt_{Ph}. *Biochemistry* **49**, 3511–3513 (2010).
9. Akyuz, N., Altman, R., Blanchard, S. C. & Boudker, O. Transport dynamics of a glutamate transporter homologue. *Nature* **502**, 114–118 (2013).
10. Erkens, G. B., Hanelt, I., Goudsmits, J. M., Slotboom, D. J. & van Oijen, A. M. Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. *Nature* **502**, 119–123 (2013).
11. Georgieva, E. R., Borbat, P. P., Ginter, C., Freed, J. H. & Boudker, O. Conformational ensemble of the sodium-coupled aspartate transporter. *Nature Struct. Mol. Biol.* **20**, 215–221 (2013).
12. Hänelt, I., Wunnicke, D., Bordignon, E., Steinhoff, H. J. & Slotboom, D. J. Conformational heterogeneity of the aspartate transporter Glt_{Ph}. *Nature Struct. Mol. Biol.* **20**, 210–214 (2013).
13. Zhao, Y. *et al.* Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188–193 (2010).
14. Stolzenberg, S., Khelashvili, G. & Weinstein, H. Structural intermediates in a model of the substrate translocation path of the bacterial glutamate transporter homologue GltPh. *J. Phys. Chem. B* **116**, 5372–5383 (2012).
15. Ryan, R. M., Kortt, N. C., Sirivanta, T. & Vandenberg, R. J. The position of an arginine residue influences substrate affinity and K$^+$ coupling in the human glutamate transporter, EAAT1. *J. Neurochem.* **114**, 565–575 (2010).
16. Borre, L. & Kanner, B. I. Coupled, but not uncoupled, fluxes in a neuronal glutamate transporter can be activated by lithium ions. *J. Biol. Chem.* **276**, 40396–40401 (2001).
17. Altman, R. B. *et al.* Enhanced photostability of cyanine fluorophores across the visible spectrum. *Nature Methods* **9**, 428–429 (2012).
18. Zheng, Q. *et al.* On the mechanisms of cyanine fluorophore photostabilization. *J. Phys. Chem. Lett.* **3**, 2200–2203 (2012).
19. Alejo, J. L., Blanchard, S. C. & Andersen, O. S. Small-molecule photostabilizing agents are modifiers of lipid bilayer properties. *Biophys. J.* **104**, 2410–2418 (2013).
20. Geggier, P. *et al.* Conformational sampling of aminoacyl-tRNA during selection on the bacterial ribosome. *J. Mol. Biol.* **399**, 576–595 (2010).
21. Grewer, C. *et al.* Individual subunits of the glutamate transporter EAAC1 homotrimer function independently of each other. *Biochemistry* **44**, 11913–11923 (2005).
22. Groeneveld, M. & Slotboom, D. J. Rigidity of the subunit interfaces of the trimeric glutamate transporter GltT during translocation. *J. Mol. Biol.* **372**, 565–570 (2007).
23. Reyes, N., Oh, S. & Boudker, O. Binding thermodynamics of a glutamate transporter homolog. *Nature Struct. Mol. Biol.* **20**, 634–640 (2013).
24. Shimamoto, K. *et al.* DL-*threo*-β-benzyloxyaspartate, a potent blocker of excitatory amino acid transporters. *Mol. Pharmacol.* **53**, 195–201 (1998).
25. Mondal, S., Khelashvili, G., Shi, L. & Weinstein, H. The cost of living in the membrane: a case study of hydrophobic mismatch for the multi-segment protein LeuT. *Chem. Phys. Lipids* **169**, 27–38 (2013).
26. Mondal, S., Khelashvili, G. & Weinstein, H. Not just an oil slick: how the energetics of protein–membrane interactions impacts the function and organization of transmembrane proteins. *Biophys. J.* **106**, 2305–2316 (2014).
27. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
28. Mitchell, P. Translocations through natural membranes. *Adv. Enzymol.* **29**, 33–87 (1967).
29. Lee, C. *et al.* A two-domain elevator mechanism for sodium/proton antiport. *Nature* **501**, 573–577 (2013).
30. Quistgaard, E. M., Low, C., Moberg, P., Tresaugues, L. & Nordlund, P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature Struct. Mol. Biol.* **20**, 766–768 (2013).
31. Zhou, X. *et al.* Structural basis of the alternating-access mechanism in a bile acid transporter. *Nature* **505**, 569–573 (2014).

Author Contributions N.A., O.B. and S.C.B. designed the study. N.A. designed and performed the majority of the experiments and simulated smFRET data. N.A. analysed the smFRET data, with support from D.S.T. and S.C.B. N.A. and O.B. analysed crystallographic data. E.R.G. performed and analysed DEER experiments and E.R.G. and J.H.F. interpreted the data. Z.Z. synthesized the 4S(COT)-maleimide cyanine dyes. D.S.T. made improvements to the smFRET instrumentation and analysis software. R.B.A. prepared reagents for smFRET experiments. H.W. and M.A.C. designed, and S.S., G.K., and M.A.C. carried out the molecular dynamics simulations. N.A., O.B. and S.C.B., H.W. and M.A.C. interpreted results and wrote the manuscript.

METHODS

Sample size. No statistical methods were used to predetermine sample size.

DNA manipulations, protein expression, purification and labelling. Single cysteine mutations were introduced by site-directed mutagenesis (Stratagene) of a cysteine-less Glt$_{Ph}$ background, in which seven non-conserved residues had been replaced with histidines resulting in improved expression levels (termed Glt$_{Ph}$ here for brevity)[3]. Constructs were verified by DNA sequencing and transformed into *E. coli* DH10-B cells (Invitrogen). Proteins were expressed as C-terminal (His)$_8$ fusions as described previously[2]. Briefly, isolated cell membranes were re-suspended in buffer A, containing 20 mM HEPES/NaOH, pH 7.4, 200 mM NaCl, 0.1 mM L-aspartate, 0.1 mM Tris(2-carboxyethyl)phosphine (TCEP). Membranes were solubilized in the presence of 40 mM *n*-dodecyl β-D-maltopyranoside (DDM) for 1 h at 4 °C. Solubilized transporters were purified by metal-affinity chromatography in buffer A supplemented with 1 mM DDM and eluted in 250 mM imidazole. The (His)$_8$-tag was cleaved by thrombin and proteins were further purified by size-exclusion chromatography (SEC). For smFRET experiments, protein samples at 40 μM were labelled with a mixture of maleimide-activated Cy3 and Cy5 dyes that exhibit enhanced photostability[17,32] as well as biotin-PEG$_{11}$, at concentrations of 50, 100 and 25 μM, respectively, for 30 min at room temperature. Labelled proteins were purified away from the excess reagents by SEC. Their purity and specificity of labelling were assessed by SDS–PAGE, which was followed by fluorescence imaging and Coomassie staining.

Protein reconstitution into liposomes for smFRET analysis and transport assays. Labelled and unlabelled Glt$_{Ph}$ variants were reconstituted into liposomes as previously described[4,11]. Briefly, liposomes, prepared from 3:1 (w/w) mixture of *E. coli* total lipid extract and egg yolk phosphotidylcholine (Avanti Polar Lipids) in a buffer containing 20 mM Tris/HEPES, pH 7.4 and 100 mM KCl, were destabilized by addition of Triton X-100 at a detergent to lipid ratio of 0.5:1 (w/w). For reconstitution, proteins were added to lipids at final protein to lipid ratio of 1:1,000 (w/w) and incubated for 30 min at room temperature. Detergents were removed by repeated incubations with Biobeads as described[11]. For smFRET and radioactive substrate uptake experiments, the same proteoliposomes were extruded through 100 nm and 400 nm filters, respectively. This reconstitution strategy yields at most 1 and 16 Glt$_{Ph}$ trimers per vesicle, respectively. Radioactive substrate uptake was measured as previously described[7]. Briefly, proteoliposomes were diluted into reaction buffer containing 20 mM Tris/HEPES, pH 7.4, 100 mM NaCl and 0.3 μM [³H]L-aspartate at room temperature. Aliquots were removed at appropriate times, diluted in ice-cold quenching buffer (20 mM Tris/HEPES, pH 7.4, 100 mM LiCl) and filtered through 0.22 μm filters (Millipore). Protein concentration was estimated by the absorbance at 280 nm after correcting for the fluorophore contributions to the value. The amount of substrate uptake was normalized per mole of Glt$_{Ph}$ monomers.

smFRET experiments. All experiments were performed using a home-built, prism-based total internal reflection fluorescence microscope constructed around a Nikon TE2000 Eclipse inverted microscope body using streptavidin-coated, passivated microfluidic imaging chambers[33]. Except when stated otherwise, labelled proteins (either detergent solubilized or liposome-reconstituted) were surface-immobilized via a biotin-streptavidin bridge. Except when stated otherwise, imaging experiments were performed in a buffer containing: 20 mM HEPES/Tris (pH 7.4), 5 mM BME, an enzymatic oxygen scavenger system comprising 1 U ml^{-1} glucose oxidase (Sigma), 8 U ml^{-1} catalase (Sigma) and 0.1% glucose[34]. In addition, apo-Glt$_{Ph}$ experiments included 200 mM KCl, Na$^+$/Asp-bound experiments included 200 mM NaCl and 0.1 mM aspartate and Na$^+$/TBOA-bound experiments included 200 mM NaCl, 10 mM TBOA. For experiments in detergent micelles, the buffers were also supplemented with 1 mM DDM. For imaging under transport conditions, the experiments were initiated in the absence of substrates (apo condition) on both sides of the membrane and chemical gradients were established by rapidly exchanging the proteoliposomes into an uptake buffer containing 100 mM NaCl and 100 μM aspartate. All data were collected at an imaging rate of 10 s^{-1} (100 ms integration time), except when otherwise stated. Fluorescence trajectories were selected for analysis using custom-made software implemented in Matlab (Mathworks) according to the following criteria[30]: a single catastrophic photobleaching event; over 8:1 signal-to-background noise ratio; a FRET lifetime of at least 5 s. FRET trajectories were calculated from the acquired intensities, I_{Cy3} and I_{Cy5}, using the formula FRET $= I_{Cy5}/(I_{Cy3} + I_{Cy5})$. Population contour plots were constructed by superimposing the FRET data from individual traces. Histograms of these population data were fit to Gaussian functions in Origin (OriginLab). The relative populations and dwell time distributions of each FRET state, as well as the transition frequencies between them, were obtained by idealizing the smFRET traces using QuB[35]. Transition density plots and the dwell time survival plots were plotted and fitted as described previously[13]. The logarithmic histograms of the dwell times were fitted to transformed probability density functions[9]. Over 300 molecules are included in each smFRET experiment to ensure that the experimental margin of error in the mean value of each distinct FRET state across the three experiments is less than 5%.

Crystallography. The R276S/M395R Glt$_{Ph}$ mutant was purified by SEC in buffer containing 10 mM Tris/HEPES, pH 7.4, 100 mM NaCl and 7 mM n-decyl-β-D-maltopyranoside (DM). Protein solution at 3.5 mg ml^{-1} was mixed at 1:1 (v:v) ratio with the reservoir solution, containing 50 mM sodium acetate, pH 5.6–6, 18–20% PEG 400 and 100–150 mM magnesium acetate, and crystallized at 4 °C by hanging-drop vapour diffusion. Crystals were cryoprotected in reservoir solution. Diffraction data were collected at National Synchrotron Light Source beamline X29. Diffraction data were indexed, integrated and scaled using the HKL2000 package[36]. Anisotropy correction was applied as described previously[5]. Further analyses were performed using CCP4 programs[37]. Initial phases were determined by molecular replacement in Phaser[38] using transport and trimeric scaffold domains as separate search models. The model was optimized by rounds of manual rebuilding in Coot[39] and refinement in Refmac5[40] with TLS[37]. During refinement, strict non-crystallographic threefold symmetry constraints were applied to the three transport domains and to regions of the scaffold domain that are involved in trimerization interactions. In addition, strict twofold symmetry constrains were applied to the entire B and C protomers, which exhibited identical positions of the transport domain. For the outward- and inward-facing states, published coordinates were used with accession numbers 2NWX (ref. 3) and 3KBC (ref. 4), respectively. For the open conformation of HP2 the accession number of the coordinates is 4OYF (ref. 6). All structural renderings were generated using PyMOL[41].

DEER measurements and data analysis. Measurements were performed at 60 K using a 17.3 GHz home-built Ku-band pulse spectrometer[42]. A standard four-pulse DEER sequence with π/2-π-π pulse widths of 16 ns, 32 ns and 32 ns, respectively, and a 32 ns π pump pulse was used routinely. The frequency separation between detection and pump pulses was 70 MHz. The detection pulses were positioned at the low-field edge of the nitroxide spectrum. The homogeneous background was removed from the raw time-domain signals and the distances were reconstructed from the baseline-corrected and normalized signals by using Tikhonov regularization method[43] and refined by maximum entropy method[44].

Molecular modelling. Molecular dynamics simulations using the Charmm27 force field (FF)[45] and updated lipid FF[46] were prepared as described previously[14] and run using the NAMD 2.9 (ref. 47) software at 300K with PME electrostatics and standard parameters for the Charmm FF. Atomic coordinates for the inward-facing wild-type Glt$_{Ph}$ were taken from PDB entry 3KBC (ref. 4) Simulations with the Gromos 54a7 FF[48] were prepared using the LAMBADA / InflateGRO membrane-embedding protocol[49] and run with the Gromacs 4.6.1 (ref. 50) simulation package with reaction-field electrostatics and standard cutoffs for the Gromos FF. All simulations included pure POPC membranes, except Charmm27 Trajectory 3 (Extended Data Fig. 9), which contained a mixture of 18% POPC, 52% POPE, and 30% POPG (prepared with Charmm-GUI web-tool[51]), more similar to the composition of the liposomes used in experiments. In selected simulations (Extended Data Fig. 9), backbone Cα atoms were subjected to harmonic restraint potentials centred on positions from the X-ray structure with a harmonic constant of 0.1 kcal mol^{-1} Å^{-2} (NAMD) or 0.24 kcal mol^{-1} Å^{-2} (Gromacs). Docking of detergent and POPC lipid molecules was performed with Autodock Vina[52] within the Chimera 1.8 visualization software[53]. Lipid insertion in Charmm27 Trajectory 3 was performed as follows: (i) a frame from the molecular dynamics trajectory after 48ns of simulation time was selected; (ii) several lipid molecules restricted to various regions of the interfaces in protomers A and C were docked, ignoring the water; (iii) docking poses among the highest ranked from all docking runs were combined, such that lipid molecules fill the available hydrophobic pockets without clashing with each other, and overlapping water molecules were discarded; (iv) local minimization was performed with the Charmm27 force field, including solvent and side chains within 5 Å of inserted lipids; (v) the molecular dynamics simulation was restarted at 300 K. Data processing and plots were performed in Matlab (Mathworks).

Kinetic simulations of smFRET data. For the simulations, we assumed that protomer motions are independent. The model presented in Fig. 6 was employed to simulate the motions of individual protomers between outward- and inward-facing orientations in QuB[35]. The time-dependent configurations of two protomers were then assigned to FRET states as described (Extended Data Fig. 3). FRET traces were generated at 100 ms time-resolution in Matlab (Mathworks) using a Gaussian distribution of FRET efficiency values and widths derived from our experimental data. Initial estimates of the kinetic parameters were based on exponential fits of the experimental dwell time distributions (Extended Data Table 1c). The parameters were then manually optimized to recapitulate the experimental observables[20]: population FRET histograms, TDPs and the dwell-time histograms (Extended Data Fig. 10).

32. Zheng, Q. *et al.* Ultra-stable organic fluorophores for single-molecule research. *Chem. Soc. Rev.* **43,** 1044–1056 (2014).
33. Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25,** 505–517 (2007).

34. Dave, R., Terry, D. S., Munro, J. B. & Blanchard, S. C. Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophys. J.* **96,** 2371–2381 (2009).

35. Qin, F. Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophys. J.* **86,** 1488–1501 (2004).

36. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

37. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).

38. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

39. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).

40. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).

41. The PyMOL Molecular Graphics System v. 1.5.0.4 (Schrödinger LLC).

42. Borbat, P. P., Crepeau, R. H. & Freed, J. H. Multifrequency two-dimensional Fourier transform ESR: an X/Ku-band spectrometer. *J. Magn. Reson.* **127,** 155–167 (1997).

43. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.* **172,** 279–295 (2005).

44. Chiang, Y. W., Borbat, P. P. & Freed, J. H. Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *J. Magn. Reson.* **177,** 184–196 (2005).

45. MacKerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126,** 698–699 (2004).

46. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114,** 7830–7843 (2010).

47. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).

48. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40,** 843–856 (2011).

49. Schmidt, T. H. & Kandt, C. LAMBADA and InflateGRO2: Efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations. *J. Chem. Inf. Model.* **52,** 2657–2669 (2012).

50. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).

51. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29,** 1859–1865 (2008).

52. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461 (2010).

53. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

54. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).

55. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
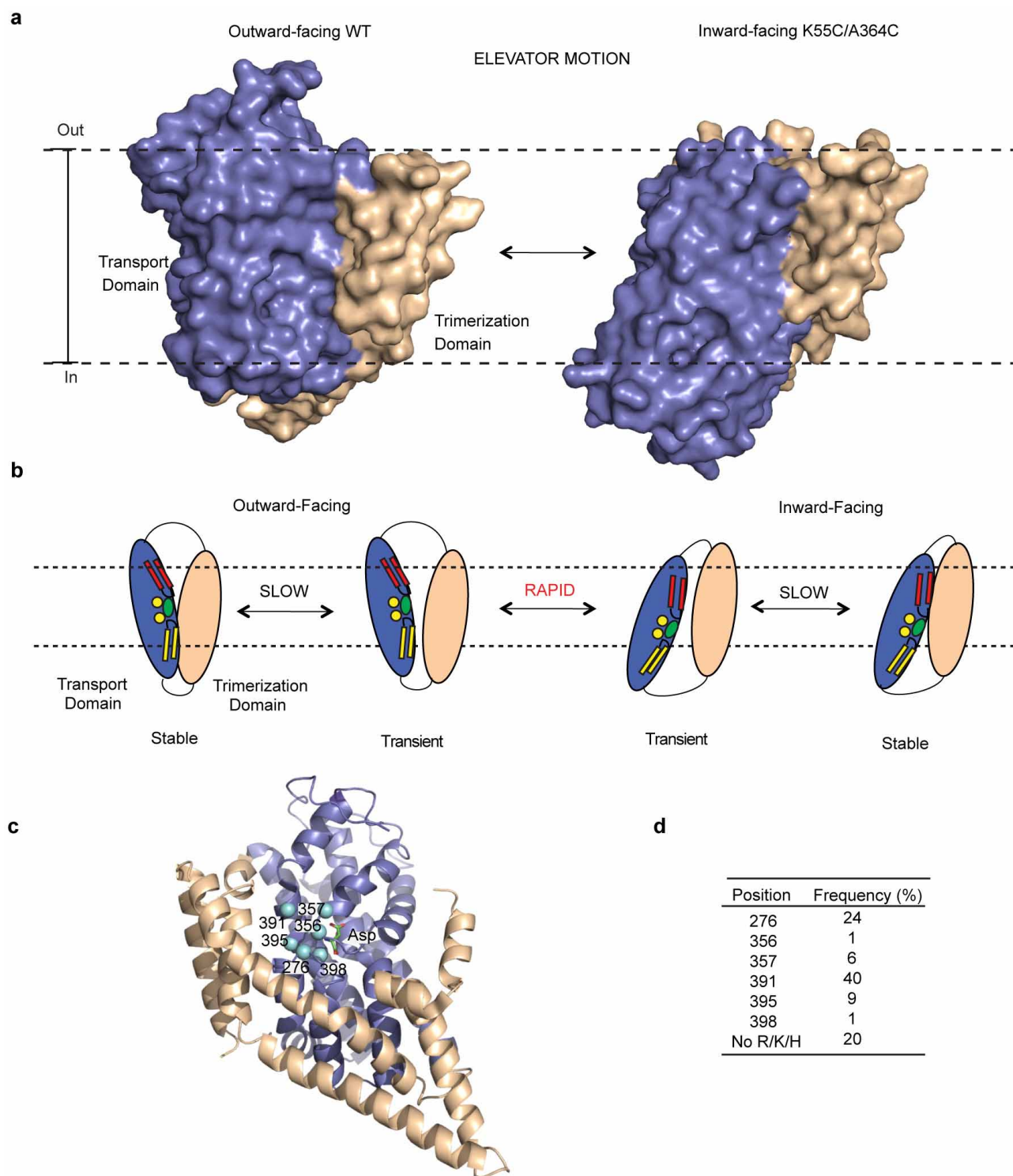
Extended Data Figure 1 | Elevator model of transport and spatial conservation of a positively charged residue in glutamate transporter family. **a**, Glt$_{Ph}$ protomers in the outward- (left) and inward-facing (right) conformation are shown in surface representation and viewed in membrane plane. Dashed lines represent an approximate position of the membrane hydrocarbon layer. In the inward-facing state, the transport domain (blue) is moved by ~15 Å across the bilayer relative to the trimerization domain (beige). **b**, Schematic representation of dynamic mode-switching between stable and transient conformations. **c**, A single Glt$_{Ph}$ protomer is shown in cartoon representation. Cyan balls emphasize the amino acid positions at which potentially positively charged residues occur in glutamate transporter homologues. **d**, Occurrence frequencies of these residues at the marked positions (Glt$_{Ph}$ numbering). To obtain the frequencies, sequences were harvested from the PFAM database[54] (accession code PF00375). Sequences were parsed to exclude those with over 70% identity and aligned using Clustal Omega[55].
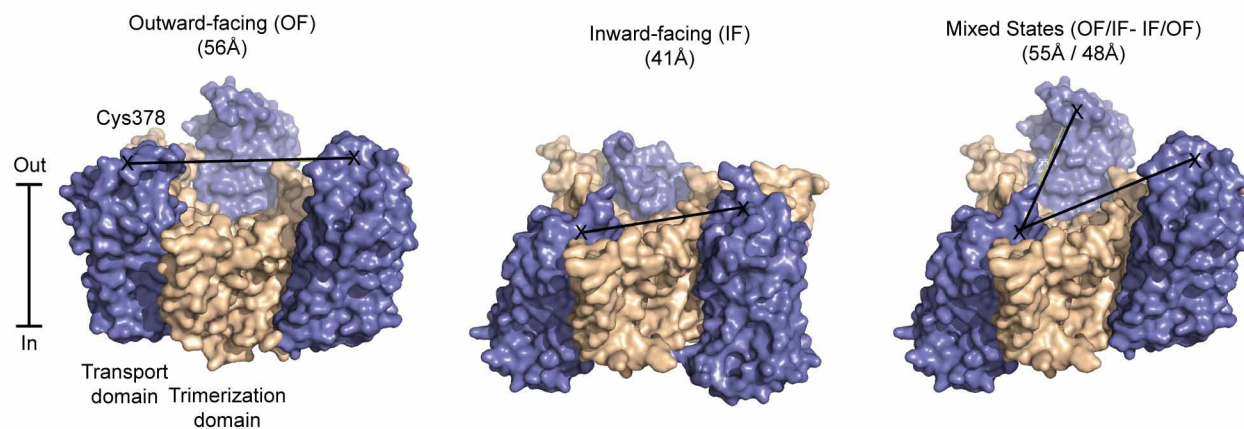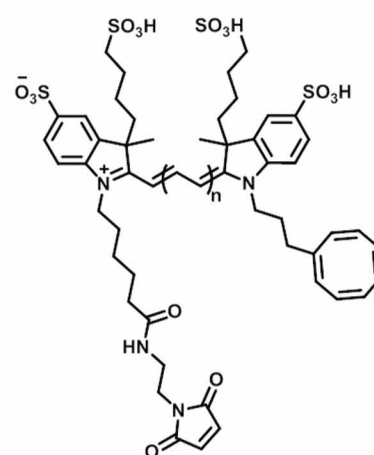
Extended Data Figure 2 | Assignment of FRET efficiency states. a, Shown are the crystal structures of Glt_{Ph} trimers in symmetrical outward (OF)- and inward (IF)-facing states and a model of an asymmetric configuration with two outward- and one inward-facing protomers[2,4]. The structures are shown in surface representation and coloured as in Extended Data Fig. 1. Black lines connect $C\alpha$ atoms of residue 378, and the corresponding distances are indicated above the structures. **b**, Expected FRET efficiency levels for these distances for all possible configurations of subunit pairs: outward/outward (OF/OF), outward/inward (OF/IF), inward/outward (IF/OF) and inward/inward (IF/IF)[4]. **c**, Intramolecularly stabilized 4S(COT)-maleimide Cy3 ($n = 1$) and Cy5 ($n = 2$) fluorophores used in this study synthesized as described previously[17,18] with the addition of two sulfonate groups for increased solubility.
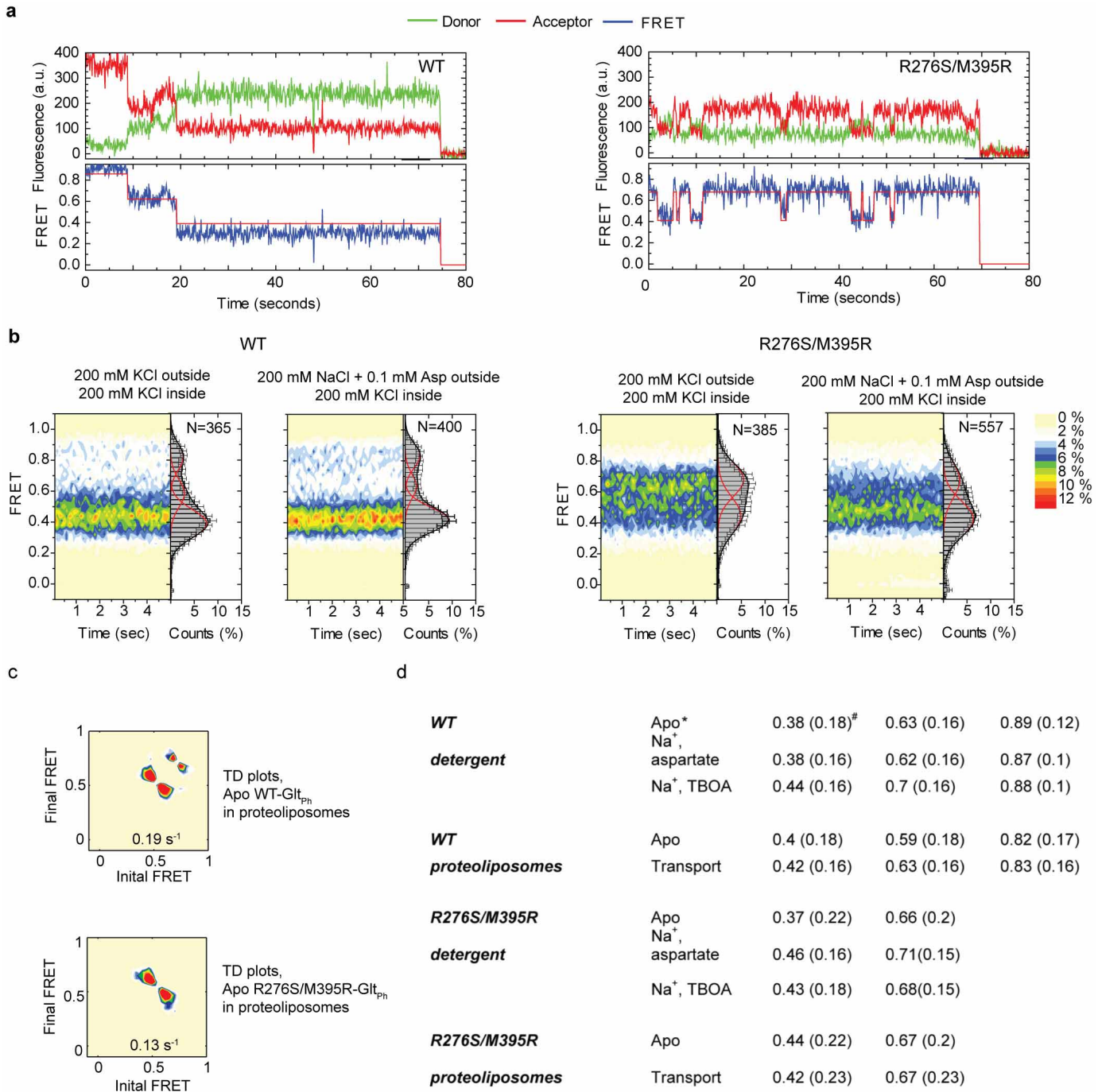
Extended Data Figure 3 | Conformational state distributions of wild-type and $H_{276,395}$-Glt_{Ph} in proteoliposomes. **a**, Examples of smFRET recordings. Top panels show raw fluorescent signals originating from donor (green) and acceptor (red) dyes. Bottom panels show changes of FRET efficiency calculated from raw data (blue). Red solid lines through the data are idealizations obtained using QuB software[35]. **b**, Contour plots and one-dimensional population histograms in the absence and presence of Na^+ and aspartate in the external liposome buffers. Buffer compositions inside and outside of the vesicles are shown above the panels. Wild-type and $H_{276,395}$-Glt_{Ph} histograms are fitted to three and two Gaussian functions, respectively. **c**, Transitions density (TD) plots for the wild type (left) and $H_{276,395}$-Glt_{Ph} (right) in proteoliposomes in the absence of Na^+ and aspartate in the external buffer. **d**, Means and widths (in brackets) of FRET efficiency distributions derived from Gaussian fits to proteoliposome data in comparison to detergent data.
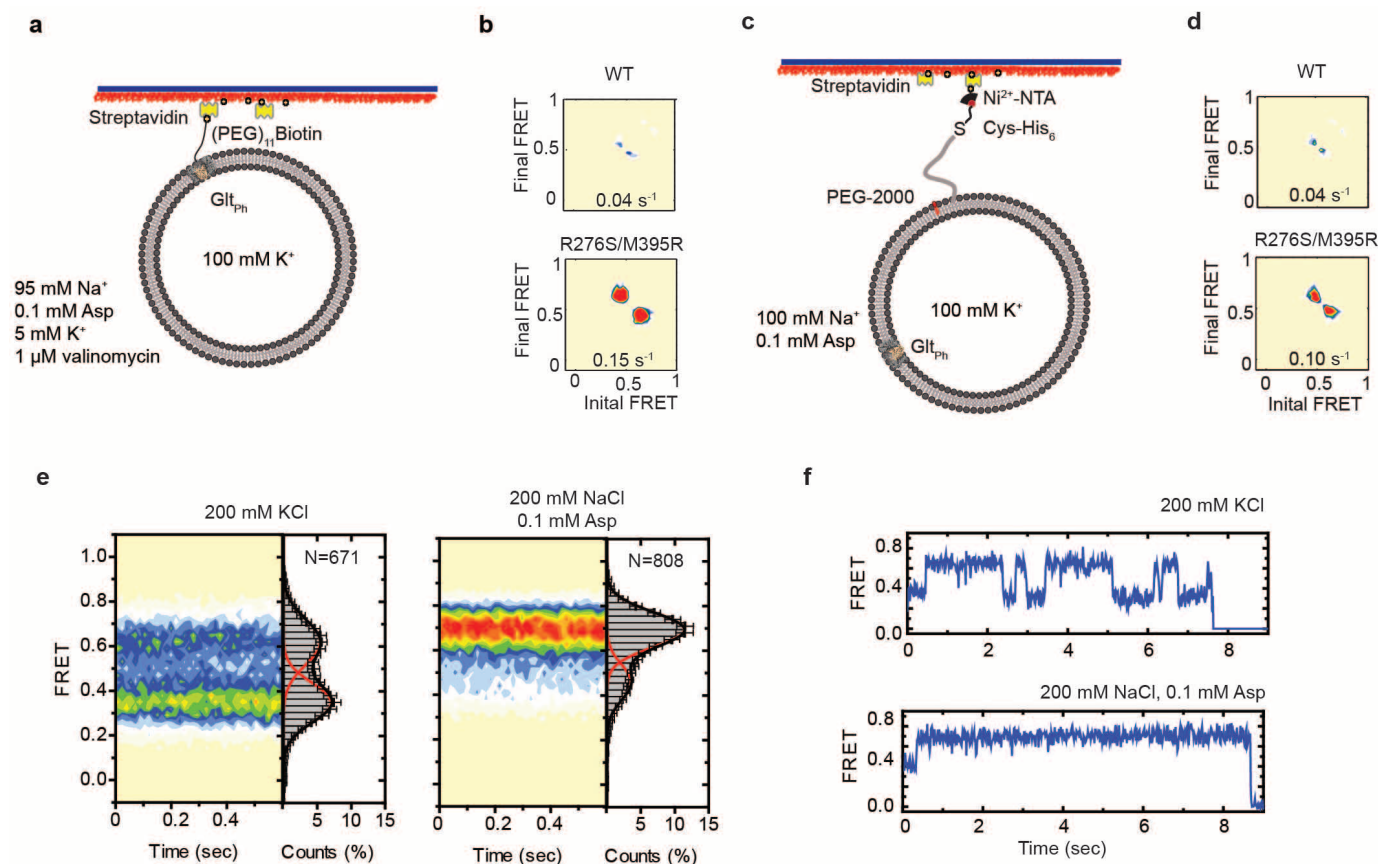
Extended Data Figure 4 | Single-molecule dynamics using different liposome-attachment strategies and with higher time-resolution.
a–d, Dynamic properties of $H_{276,395}$-Glt_{Ph} under transport conditions using a different surface-immobilization strategy and in the presence of electrical potential. **a,** Surface-immobilization strategy for proteoliposomes using His-tagged lipids. **b,** Transition frequencies for wild-type (top) and $H_{276,395}$-Glt_{Ph} (bottom) trimers reconstituted into his-tagged liposomes that were site-specifically labelled in just two protomers with intramolecularly photostabilized Cy3 and Cy5 fluorophores. **c,** A negative inside voltage potential was established in proteoliposomes by adding valinomycin to the

uptake buffer. **d,** Transition frequencies for wild-type (top) and $H_{276,395}$-Glt_{Ph} (bottom) in the presence of valinomycin. Each experiment shown includes statistics based on >250 individual molecules. The standard error in transition frequency measurements is approximately $0.015\ s^{-1}$. **e, f,** Dynamic properties of $H_{276,395}$-Glt_{Ph} probed at 15 ms time resolution. Contour plots and one-dimensional population FRET efficiency histograms (**e**) observed for the humanized mutant in detergent solution in the absence (left) and presence (right) of 100 mM NaCl and 100 μM aspartate. Examples of single-molecule trajectories are shown in **f**.
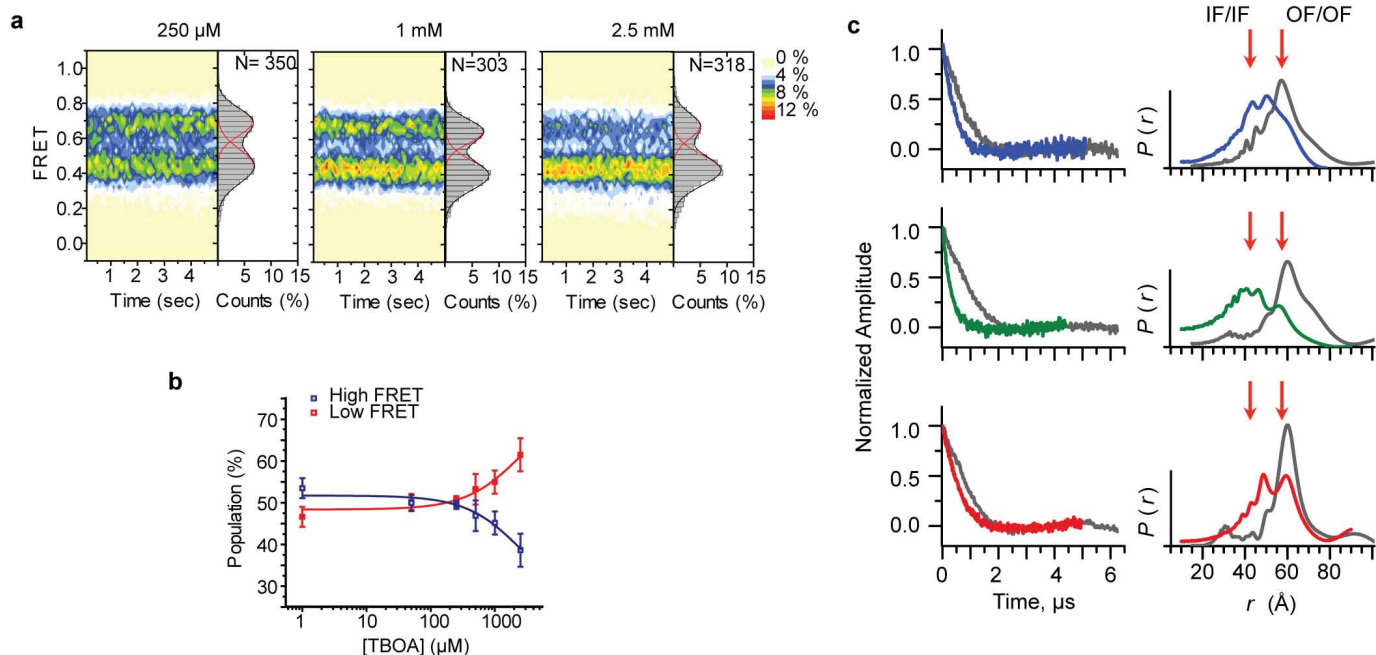
Extended Data Figure 5 | Population changes in response to ligand binding. a, b, TBOA binding to $H_{276,395}$-Glt_{Ph} measured in smFRET experiments. Contour plots and population FRET efficiency histograms in the presence of increasing concentrations of TBOA (**a**). Changes in low- (red) and high- (blue) FRET state populations as a function of TBOA concentration (**b**). Solid lines through the data correspond to the Hill equation $y = y_{min} + (y_{max} - y_{min})$ $(x^n/(x^n + K_d^n))$ with $K_d = 2.4$ mM and $n = 1$. The data points shown are averages and standard errors from three independent biological replicates. **c**, Experimental time domain DEER data (left) and reconstructed distance distributions (right) for $H_{276,395}$-Glt_{Ph} (shown in colours) and wild-type transporter (black) spin-labelled on residue Cys378 in detergent solution. The data were collected in the absence of ligands (top), in the presence of 100 mM Na^+ and 350 μM aspartate (middle) and in the presence of 100 mM Na^+ and 480 μM TBOA (bottom). The red arrows above the distance distributions mark distances between residues 378 extracted from crystal structures of the symmetric outward- (OF/OF) and inward- (IF/IF) facing states. The data for the wild-type transporter were adapted from a published study[11]. The data show that in the apo transporter, outward- and inward-facing states are similarly populated. Binding of Na^+ ions and aspartate favours the inward-facing state, whereas binding of TBOA favours the outward-facing state.
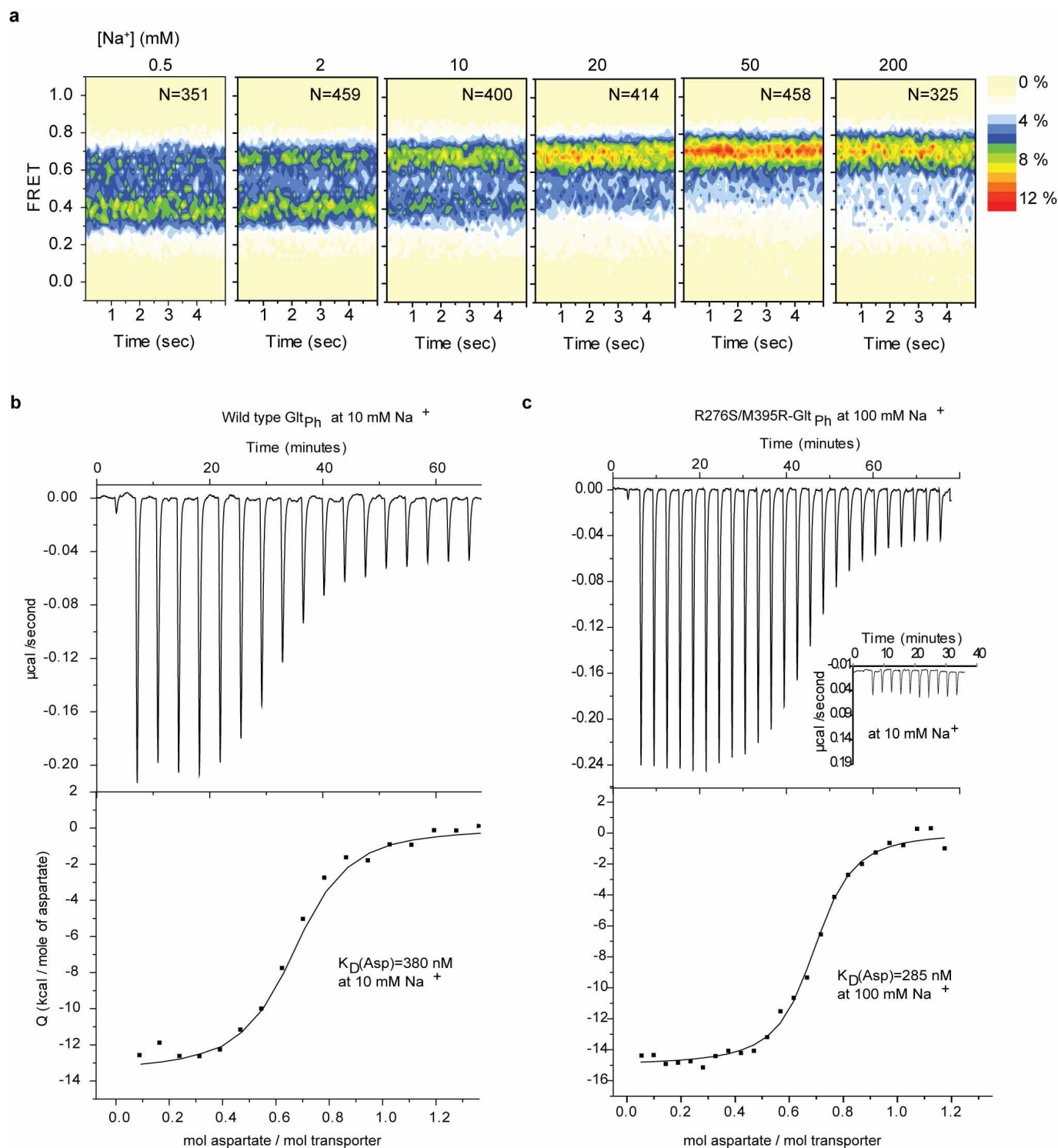
Extended Data Figure 6 | Aspartate binding experiments. a, FRET efficiency population contour plots determined for $H_{276,395}$-Glt_{Ph} in detergent micelles in the presence of 100 μM aspartate and increasing concentrations of Na^+ ions (indicated above the panels). **b, c**, Representative aspartate binding isotherms derived from ITC experiments for the wild-type Glt_{Ph} (**b**) and $H_{276,395}$-Glt_{Ph} (**c**) in the presence of 10 mM Na^+ and 100 mM Na^+, respectively. The binding of aspartate to $H_{276,395}$-Glt_{Ph} in the presence of 10 mM Na^+ is too weak to measure (inset). Binding experiments were performed using small-volume Nano ITC (TA Instruments). Upper panels show raw data. The cell contained 30 μM (WT-Glt_{Ph}) and 40 μM ($H_{276,395}$-Glt_{Ph}) protein buffer containing 20 mM HEPES/Tris, pH 7.4 and 0.1 mM DDM and indicated concentrations of NaCl. The syringe contained Asp at 200 μM concentration in the same buffer; every injection contained 5 μl. Data were processed and analysed using manufacturer's software (lower panels). Solid lines through the data are fits to independent binding sites model with the following K_d, enthalpy (ΔH), and apparent number of binding sites (n): 380 nM, 15 kcal per mol and 0.65 for the wild-type transporter, and 285 nM, 16 kcal per mol and 0.68 for $H_{276,395}$-Glt_{Ph}.

a

Data Collection

Space group	P6522
Cell a, c (Å)	122 122 580
Cell α, β, γ (°)	90, 90, 120
Wavelength, nm	1.075
Beamline	NSLS X29

Scaling

Resolution (Å)	50-4.2*
Rsym (%)	9 (N/A)
I/ σ	23 (2.58)
Completeness (%)	82 (15)
Redundancy	10.8 (16.9)

Refinement

Resolution (Å)	20-4.2
No. reflections	15417
Rwork/ Rfree	28/31
No. atoms	9097
Mean B-factors (Å2)	132
Bond angles (°)	1.1272
Bond length (Å)	0.0077
Ramachandran (%)	
Favored	85.4%
Outliers	4.1%

b

c

d

Locked protomer A

Unlocked protomers B and C

Extended Data Figure 7 | Data collection and refinement for Na$^+$ and aspartate bound H$_{276,395}$-Glt$_{Ph}$. **a**, Table showing data collection and refinement statistics. Scaling and refinement statistics were obtained after anisotropy correction by ellipsoidal truncation using high-resolution cutoffs of 4.9 Å along the *a* and *b* axis, and of 4.2 Å along the *c* axis. **b**, Stereoview of the $2F_o–F_c$ electron density map for H$_{276,395}$-Glt$_{Ph}$ contoured at 1.5 σ around residue Arg 395 in unlocked protomer C. Protein backbone (maroon) is shown in cartoon representations and side chains are shown as lines and colored by atom type. **c**, Superimposed scaffold domains of the inward-facing wild type and H$_{276,395}$-Glt$_{Ph}$ are shown in cartoon representation. The labile portions are coloured cyan (wild type) and magenta (mutant). Helices bend at conserved Pro 60 and Pro 206 residues (spheres). **d**, Locked (left) and unlocked (right) mutant protomers viewed from the cytoplasm and shown in surface representation.
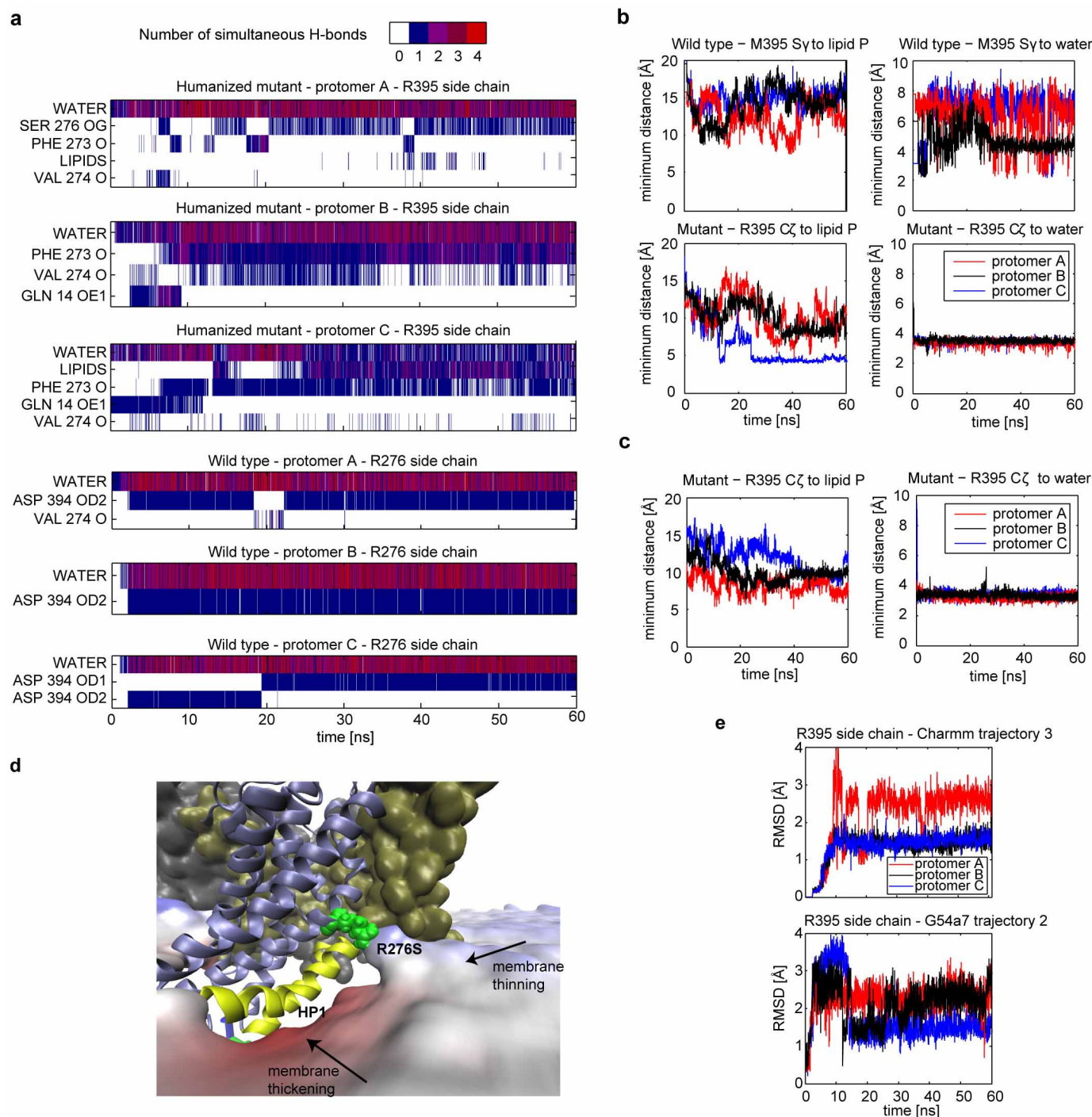
Extended Data Figure 8 | Arg395 adapts to its environment. a, The arginine side chain (Arg 276 in the wild type; Arg 395 in $H_{276,395}$-Glt_{Ph}) is seen in molecular dynamics simulations to engage in hydrogen-bonding interactions. The extent of the hydrogen bonds formation is shown as a function of simulation time in Charmm Trajectory 3 (see Extended Data Fig. 10). The main interactions of the arginine in both mutant and wild type are with water molecules, but the locations of the waters are very different. In $H_{276,395}$-Glt_{Ph}, the Arg 395 side chain is located 5 to 9 Å below the level of the membrane surface, so that the water molecules are those penetrating the membrane–protein interface due to remodelling of the membrane. In the wild type, the water molecules interacting with Arg 276 are in the space created inside the protein. **b**, The minimum distance from wild-type Met395 (top) or mutant Arg 395 (bottom) side chains to any lipid phosphate group (left) or any water molecule (right) in Charmm Trajectory 3. In $H_{276,395}$-Glt_{Ph}, after the initial equilibration phase, lipid phosphate groups interact with Arg 395 either directly (5 Å distance) or through water (7.5 Å distance). In the wild type, lipid

head groups remain far from the hydrophobic Met 395 side chain. Water interacts constantly with Arg 395, but only occasionally with Met 395 (in protomer B, a water molecule approaches Met 395 from the inside of the protein, at the interface between transport and trimerization domains). **c**, The same set of distances as in **b** for the mutant, from a different trajectory (G54a7 Trajectory 2) obtained independently, using a different force field. The same trends are observed as in **b**, showing proximity to the polar environment. **d**, Membrane bending (blue indicates thinning, red indicates thickening) close to Arg 395 (green) which exposes its side chain to a polar environment comprised of water molecules and lipid head groups. **e**, Root mean square deviation (r.m.s.d.) of the Arg 395 side chain with respect to the crystal structure after alignment on the trimerization domain, calculated from Charmm Trajectory 3 and G54a7 Trajectory 2. The side chain initially samples different conformations before settling into the membrane-exposed position shown in panel **d**.
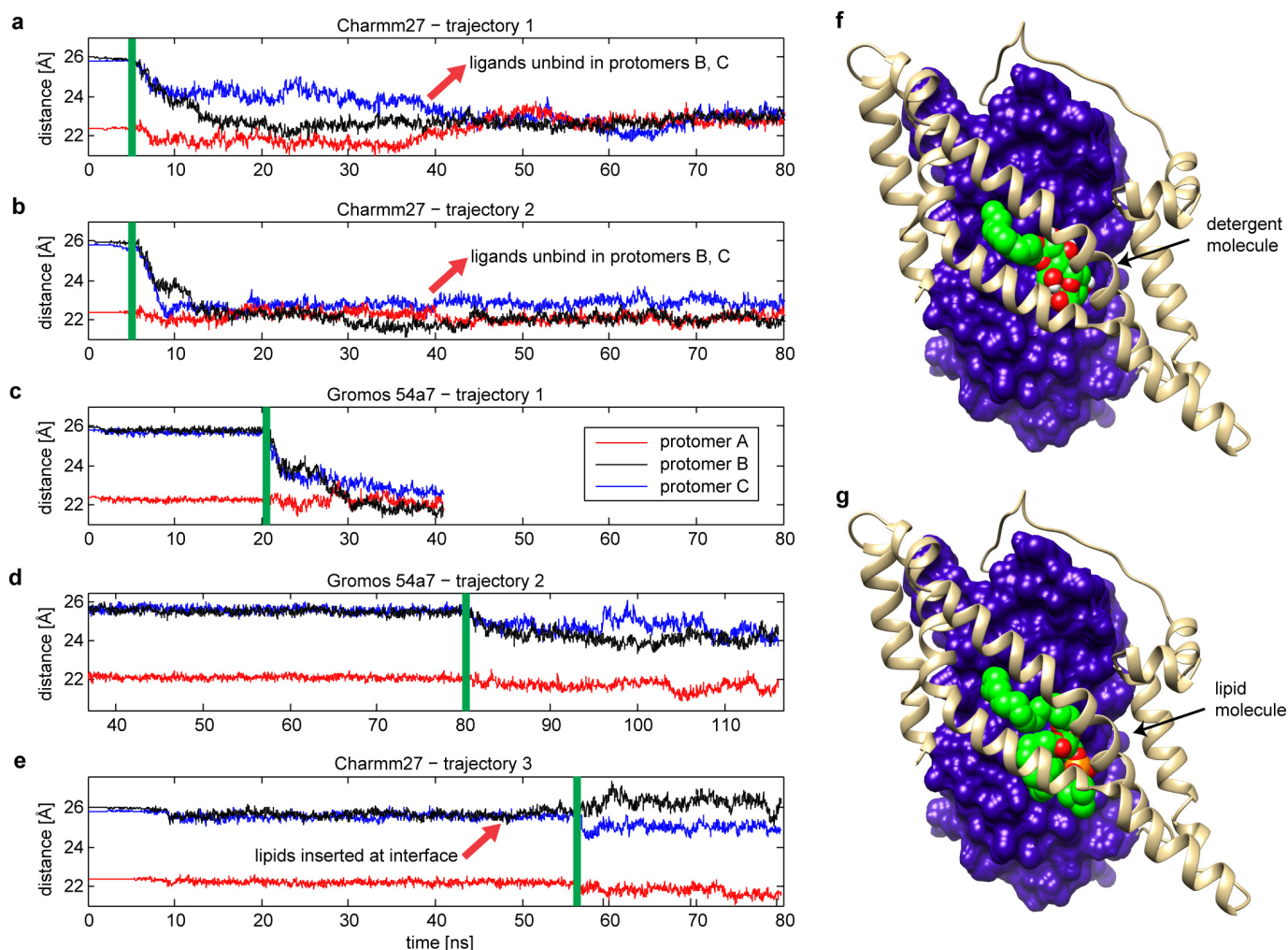
Extended Data Figure 9 | Lipids or detergent molecules stabilize the unlocked conformation of H$_{276,395}$-Glt$_{Ph}$. **a–e,** Centre-of-mass distance between the transport and scaffold domains of protomers A, B, and C of H$_{276,395}$-Glt$_{Ph}$ as a function of molecular dynamics simulation time. The data are from five independent simulations initiated with position restraints on the Cα atoms (later released at different time points) and with the domain interface solvated with water. The vertical green lines indicate the moment in the corresponding trajectory when position restraints were turned off. Panels **a** and **b** show two repeats of the same starting structure simulated with the Charmm force field[45] and panel **c** with Gromos force field[48]. The transport domains in protomers B and C collapse onto the trimerization domain rapidly and lose their ligands in some cases (red arrows). **d,** A simulation, in which lipid tails partially insert into the interface spontaneously; the unlocked structure is stable much longer (note the different time scales on the time axis), and the collapse is only partial. **e,** The trajectory of a NAMD simulation (Charmm force field) in which lipid molecules were docked into the interface of protomers B and C at the time marked by the red arrow (3 lipids per protomer). The lipids remained in the docked region for the entire duration of the simulation and stabilized the position of the transport domain. **f, g,** The best scored docking poses for a detergent molecule and a POPC lipid, respectively, docked at the interface of protomer C.
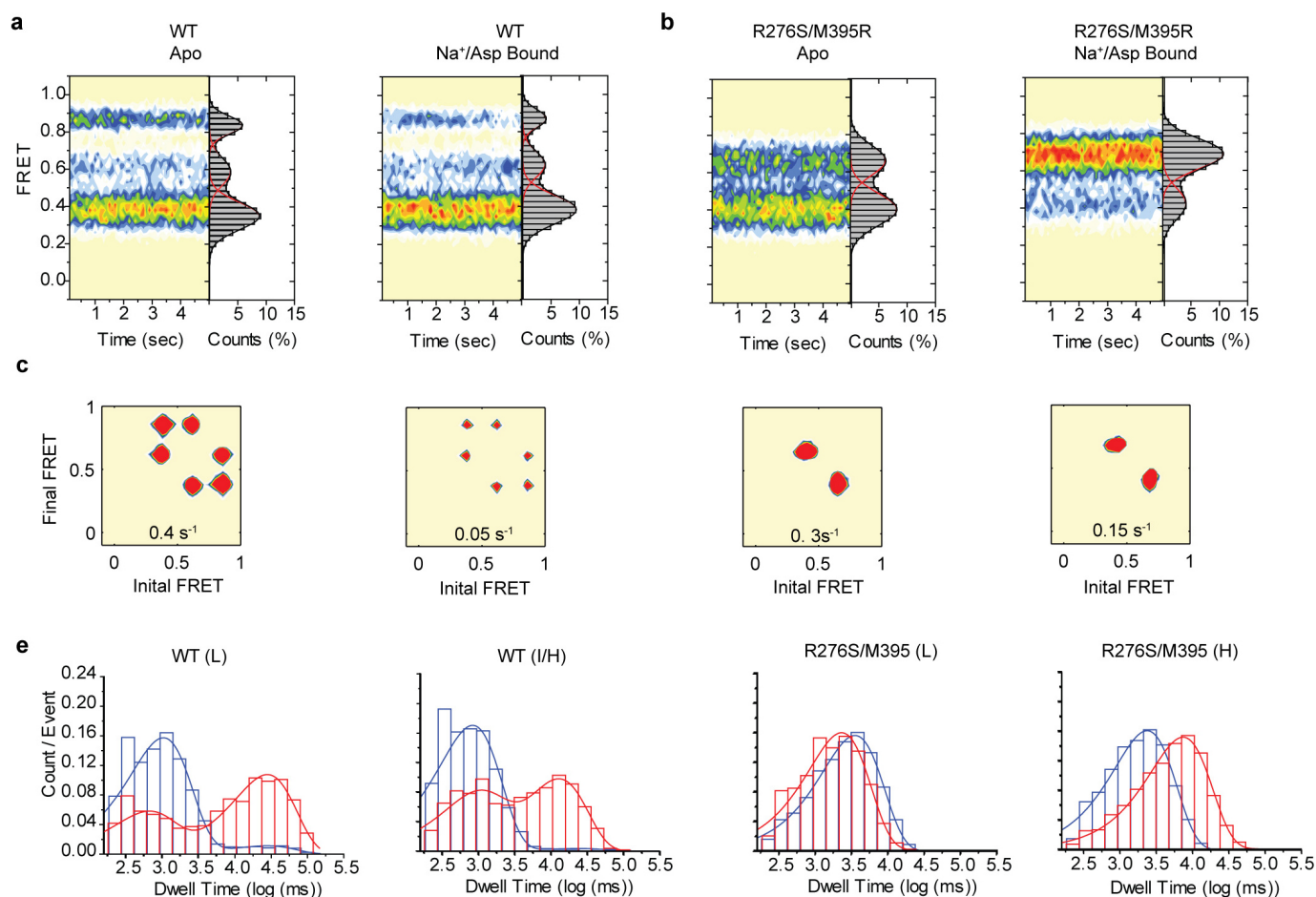
Extended Data Figure 10 | Simulated smFRET data recapitulate experimental observations. a–d, Simulated FRET efficiency population contour plots (left side of each panel) and cumulative population histograms (right side) for wild-type Glt$_{Ph}$ (a) and H$_{276,395}$-Glt$_{Ph}$ (b), and the corresponding transition density plots (c and d), (see Fig. 2 for corresponding experimental data). As noted before[9], there are fewer transitions observed between the low- and high-FRET states in the wild-type transporter than would be expected from the model. This may either be because the model does not recapitulate the noise correctly or it may reflect previously uncharacterized communication between the protomers that warrants further investigation. e, f, Dwell time distributions for the low- (left panel) and intermediate- and high-FRET states (right panels) obtained for wild-type Glt$_{Ph}$ (e) and H$_{276,395}$-Glt$_{Ph}$ (f) (see Fig. 4 for corresponding experimental data).

Extended Data Table 1 | FRET state assignments and populations; time constants for the slow and fast components

a, FRET State Population Distributions in proteoliposomes

WT Glt$_{Ph}$

FRET	Subunit configuration	Apo, %	P(out)=0.55, %	Transport, %	P(out)=0.65, %
Low	OF/OF+OF/IF	54	55	63	65
Intermediate	IF/OF	27	25	22	22
High	IF/IF	19	20	15	13

R276S/M395R Glt$_{Ph}$

FRET	Subunit configuration	Apo, %	Transport, %
Lower	OF/OF+OF/IF	40	55
Higher	IF/OF+IF/IF	60	45

b, FRET State Population Distributions in detergent micelles

WT Glt$_{Ph}$

FRET	Subunit configuration	Apo, %	P(out)=0.45, %	Bound, %	P(out)=0.5, %
Low	OF/OF+OF/IF	46	45	49	50
Intermediate	IF/OF	24	25	25	25
High	IF/IF	29	30	24	25

R276S/M395R Glt$_{Ph}$

FRET	Subunit configuration	Apo, %	Transport, %
Lower	OF/OF+OF/IF	62	30
Higher	IF/OF+IF/IF	38	70

c, Time constant for stable (slow) and transient (fast) FRET States in detergent micelles

WT	Low FRET		Intermediate / High FRET	
	t_{Fast}, s	t_{Slow}, s	t_{Fast}, s	t_{Slow}, s
Apo	~ 0.6	~ 6	~ 0.6	~ 5
Na$^+$, aspartate	~ 0.7	~12	~ 0.7	~ 15
R276S/M395R	**Low FRET**		**Higher FRET**	
	t, s		t, s	
Apo	~ 1.5		~ 1.1	
Na$^+$, aspartate	~ 1.7		~ 7.6	

a, **b**, Shown are the assignments of FRET states to configurations of labelled subunit pairs and corresponding observed populations, rounded to integer numbers. Also shown are the calculated populations considering the probability of a protomer to be in the outward facing state *P(out)* and assuming independent protomers in the trimer. **c**, Time constants for the wild-type transporter, τ, of the slow and fast components were derived from fitting the survival data compiled from the measured dwell times to double exponential function. The time constants for the H$_{276,395}$-Glt$_{Ph}$ mutant were obtained by fitting the survival data to a single exponential function. Shown are averages from three independent experiments. The standard errors are within 5%. Dwell times longer than 10 s are significantly underestimated because photobleaching, which occurs with time constant of ~40 s, is limiting the observation window.

LETTER

A possible close supermassive black–hole binary in a quasar with optical periodicity

Matthew J. Graham[1], S. G. Djorgovski[1], Daniel Stern[2], Eilat Glikman[3], Andrew J. Drake[1], Ashish A. Mahabal[1], Ciro Donalek[1], Steve Larson[4] & Eric Christensen[4]

Quasars have long been known to be variable sources at all wavelengths. Their optical variability is stochastic and can be due to a variety of physical mechanisms; it is also well-described statistically in terms of a damped random walk model[1]. The recent availability of large collections of astronomical time series of flux measurements (light curves[2–5]) offers new data sets for a systematic exploration of quasar variability. Here we report the detection of a strong, smooth periodic signal in the optical variability of the quasar PG 1302−102 with a mean observed period of 1,884 ± 88 days. It was identified in a search for periodic variability in a data set of light curves for 247,000 known, spectroscopically confirmed quasars with a temporal baseline of about 9 years. Although the interpretation of this phenomenon is still uncertain, the most plausible mechanisms involve a binary system of two supermassive black holes with a subparsec separation. Such systems are an expected consequence of galaxy mergers and can provide important constraints on models of galaxy formation and evolution.

Subparsec supermassive black-hole (SMBH) binary systems are not resolvable except possibly with long baseline radio interferometry. An alternative approach to their detection is through a modulated variability—caused by, for example, perturbations in their accretion disks or precession of relativistic jets, if they are present (Fig. 1). The best known candidate, OJ 287[6], has shown a pair of outburst peaks every 12.2 years for at least the past century: this object can be interpreted as a secondary SMBH perturbing the accretion disk of the primary SMBH at regular intervals[7]. Systematic searches for equivalent systems to date[8–10] have attempted to identify them from broad-line velocity offsets in their optical and near-infrared spectra, but cannot detect the closest pairs (with a separation of ≲ 0.1 pc).

We applied a novel joint wavelet and autocorrelation function (ACF) based technique that identifies objects exhibiting strongly periodic behaviour in their light curves (M.J.G. *et al.*, manuscript in preparation) to the largest set of quasar time series currently available. These are drawn from the Catalina Real-time Transient Survey (CRTS)[11–13]. PG 1302−102 (Fig. 2) is the strongest periodic candidate out of 20 objects meeting the selection criteria: strong constant wavelet peak, strong ACF detection of periodic behaviour, sufficient temporal coverage for 1.5 or greater cycles at the detected period, and a phased light curve well-described by a sinusoid. For statistical comparison, we have also generated a simulated light curve for each known quasar based on a damped random walk (DRW) model, a standard statistical description of the optical variability of quasars[1], using the CRTS time sampling. We find that only one object from the simulated data of 247,000 quasars satisfies the same selection criteria, showing that the number of quasars selected is statistically significant and that strongly periodic behaviour is not expected as an artefact of a DRW process.

PG 1302−102 has a median V-band magnitude of 15.0 and a redshift of 0.2784, which gives an absolute V-band magnitude of $M_V = -25.81$, assuming the 9-year WMAP cosmology[14]. It is outside the footprint of the Sloan Digital Sky Survey but is associated with bright infrared and X-ray sources. It is also a very bright (720 mJy at 4.86 GHz), core-dominated flat spectrum radio source. Its optical/near-infrared spectrum (Fig. 3) shows broad emission lines (Hβ, Hα, Paβ, Paα) with an inferred mass of $\log(M/M_\odot) = 8.3$–9.4 and the object appears to be radiating at or close to its theoretical Eddington limit ($\log(L/L_{Edd}) = 0$). The light curve for the quasar is well-fitted by a sinusoid with an observed period of 1,884 ± 88 days (corresponding to a rest-frame period of 1,474 ± 69 days) and an amplitude of ∼0.14 mag. CRTS data (covering ∼1.8 cycles; that is, May 2005 to the present day) are augmented by archival monitoring data[15,16] available back to May 1993, giving a total coverage of 4.1 cycles. These data are consistent with the behaviour seen in the past nine years of CRTS data, particularly with stochastic photometric variation imposed on a periodic signal. Further simulations show that the detection is statistically significant, with an observed signal 40 times the scatter from the mean.

As PG 1302−102 is bright and nearby, it has featured in a number of studies of quasars and their host galaxies. The radio and optical structure of the source is noted to be unusual. Hubble Space Telescope (HST) imaging[17] shows that the quasar resides in a luminous elliptical host, as typical for radio-loud quasars[18]. There are also two companion galaxies that lie at projected distances of 3 and 6 kpc. Several features in radio
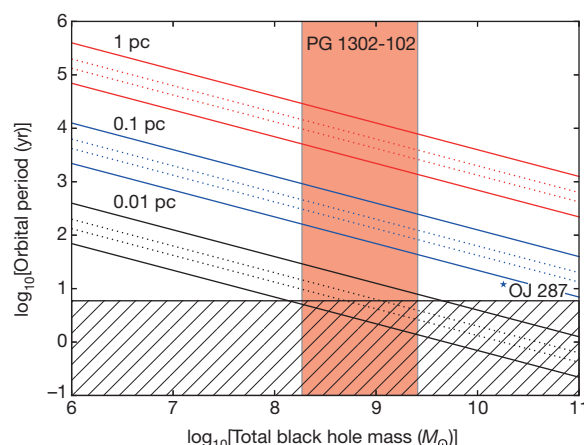
Figure 1 | The parameter space of SMBH binary pairs. The expected orbital periods for SMBH close binary pairs at the specified separations as a function of total black-hole mass. The solid upper line for each separation indicates a $z = 5$ track and the solid lower line a $z = 0.05$ track, while the two internal dotted lines show $z = 1.0$ (lower) and $z = 2.0$ (upper) tracks, respectively. The hatched region indicates the range over which CRTS has temporal coverage of 1.5 cycles or more of a periodic signal. The pink shaded region shows the region of detection for the best CRTS candidate given the range of virial black-hole masses reported in the literature. Also shown (solid black star) is the location of the best known SMBH binary candidate, OJ 287[6].

[1]California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [3]Department of Physics, Middlebury College, Middlebury, Vermont 05753, USA. [4]Lunar and Planetary Laboratory, Department of Planetary Sciences, University of Arizona, 1629 East University Boulevard, Tucson, Arizona 85721, USA.
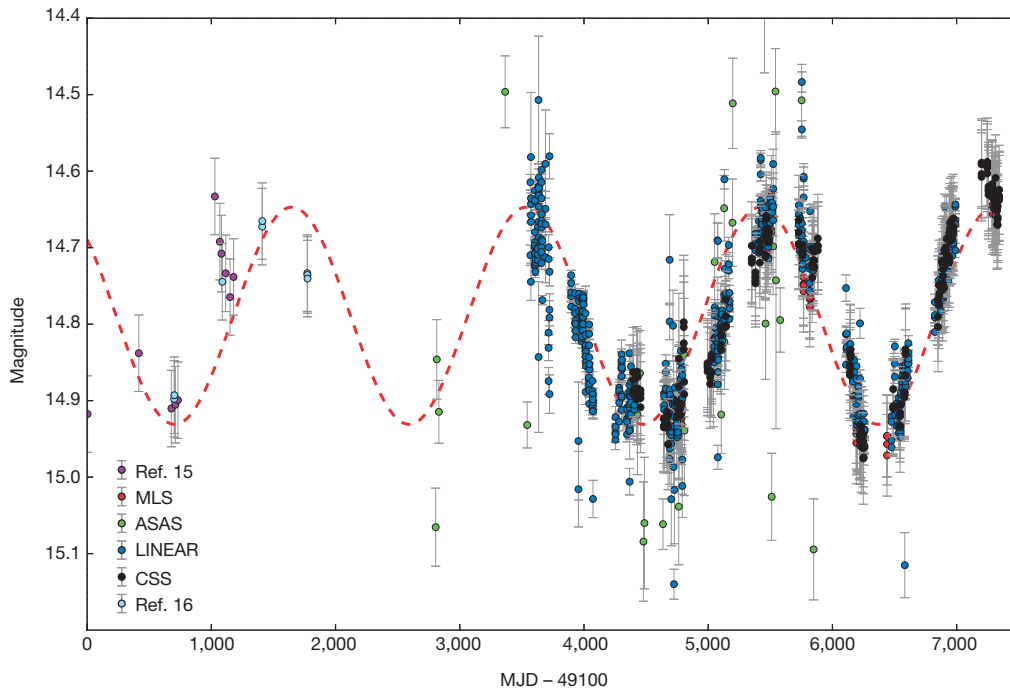
Figure 2 | The composite light curve for PG 1302−102 over a period of 7,338 days (∼20 years). The light curve combines data from two CRTS telescopes (CSS and MLS) with historical data from the LINEAR and ASAS surveys, and the literature[15,16] (see Methods for details). The error bars represent one standard deviation errors on the photometry values. The red dashed line indicates a sinusoid with period 1,884 days and amplitude 0.14 mag. The uncertainty in the measured period is 88 days. Note that this does not reflect the expected shape of the periodic waveform, which will depend on the physical properties of the system. MJD, modified Julian day.

images of PG 1302−102, such as the small radio core and sharp bends in the radio structure very close to the central source, correspond with features seen in the optical[19]. An interpretation is that the host galaxy is a fairly old merger, but that there might be more recent activity, with the radio source just turning on and possible radio jets just emerging from the host galaxy. There may also be some indication of relativistic beaming connected with a jet. It should be noted that OJ 287 also exhibits a similar radio and optical morphology[20].

PG 1302−102 was spectroscopically monitored over a six-month period in 1990[21] and showed no detectable (greater than 5–10%) change in any component of its spectrum over that time. This lack of variation is not inconsistent with the ∼60 month period that we have identified. A SMBH binary may also exhibit double-peak broad line profiles in its spectrum for a small window of separation between the pair[22] (although disk emission from an accretion disk around a single source may also produce the same effect[23]). At closer distances, the two black holes dynamically affect the broad-line region clouds as a single complex entity producing single-peaked spectral lines with asymmetric line profiles. The Balmer

and Paschen series spectral lines in PG 1302−102 do not show a double peak profile but are consistently asymmetric (Fig. 4). In particular, a small bump on the red wing of Hβ has been reported[21], implying a velocity shift of the order of 200 km s^{-1} between the narrow and broad components of Hβ. One proposed explanation for this is a binary system.

Figure 4 | The profiles of the Balmer and Paschen series lines of PG 1302−102. The data have been modelled with a multi-component line fitting technique (see Methods for details). **a, b,** Balmer Hβ (**a**) and Hα (**b**) have been fitted using a narrow component (dashed blue line) and a broad Gaussian (solid orange line). The dashed green line shows the linear continuum component, and the total fitted profile is shown as a solid red line. Hβ requires a single Gaussian offset from the narrow component but Hα requires two components—a central Gaussian plus a red wing. **c, d,** The Paschen lines (Paβ (**c**) and Paα (**d**)) also show a consistent small asymmetry on the red side.

Figure 3 | The composite spectrum for PG 1302−102. This combines an archival GALEX spectrum (ultraviolet) with optical/near-infrared spectra taken with the Keck and Palomar 200 inch telescopes in April and June 2014. F_λ, flux density. The prominent emission lines are indicated. The median flux errors are 5.6×10^{-16} erg s^{-1} cm^{-2} for the GALEX data ($\lambda < 0.3$ μm), 4.5×10^{-16} and 7.6×10^{-17} erg s^{-1} cm^{-2}, respectively, for the blue (0.3 μm $< \lambda < 0.5$ μm) and red (0.5 μm $< \lambda < 0.9$ μm) optical spectra from Palomar, and 4.6×10^{-18} erg s^{-1} cm^{-2} for the Keck near-infrared ($\lambda > 0.9$ μm) spectrum.

The physical interpretation of the periodicity is uncertain, although its sinusoidal nature suggests that it is kinematic in origin: we consider three possibilities here. (1) The optical flux could be the superposition of thermal emission from the accretion disk and a non-thermal contribution from a precessing jet, and such a model can fit the observed data (see Methods). The expected precession period with a single SMBH is about $10^{2.8}$–$10^{6.5}$ years[24], much longer than the observed period. Thus, a binary SMBH origin for the jet precession is more plausible. In this latter case, the jet could precess for two reasons: either as a result of inner disk precession due to the tidal interaction of an inclined secondary SMBH, or because of the precession of a circumbinary disk warped by the SMBH binary. (2) Another possibility is a temporary hotspot in the inner region of an accretion disk, but this leads to implausible single SMBH mass estimates of $\log(M/M_{\odot}) = 11.4$–12.2, depending on the degree of rotation of the SMBH (the largest reported SMBH masses[25] are of the order of $\log(M/M_{\odot}) \approx 10$). However, with a SMBH binary, periodic mass accretion rates can give rise to an overdense lump in the inner circumbinary accretion disk[26]. The spectral energy distribution of a circumbinary disk also has a steeper power law[27] and so accretion variations will have a more noticeable effect at shorter wavelengths. (3) Yet another possibility is a warped disk eclipsing part of the continuum as it precesses, although SMBH binaries are proposed as a possible cause for such warped disks[28]. We note as well that light curves for objects known to exhibit these phenomena do not resemble that of PG 1302−102 (see Extended Data).

If PG 1302−102 were to be described as a binary SMBH pair with a total virial mass of $\log(M/M_{\odot}) \approx 8.5$, the observed period gives an upper-limit separation of ~0.01 pc between the pair. This would mean that the system has evolved well into the 'final parsec' scale. The expectation is that most binary SMBH systems will spend the majority of their lifetime at such separations (0.01–1 pc), in an intermediate phase of evolution between scattering any stars in the nuclear region and gravitational radiation dominance[29].

Further observations could test the different interpretations mentioned above, particularly reverberation mapping to measure the behaviour of emission line response to continuum variations, which is expected to be different for different explanations[22]. Continued monitoring by CRTS and other synoptic surveys will track future cycles, and historical photometric data from photographic plate collections may provide more data for previous ones. With decadal baselines, the predicted change in period of the system may be detectable. Future spectroscopic observations could also test whether the line asymmetries vary on binary orbital timescales. Multiwavelength observations should provide more information about the innermost regions of the quasar and the nature of the jet. The relationship between PG 1302−102 and its two nearby companions may also furnish insight into the merger history of this source, particularly as these may contain similarly sized SMBHs. Finally, if PG 1302−102 is a SMBH binary, it is a strong candidate for any gravitational wave experiment sensitive to nanohertz frequency waves, such as those using pulsar timing arrays and any future space-borne gravitational wave detection mission.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Kelly, B. C., Bechtold, J. & Siemiginowska, A. Are the variations in quasar optical flux driven by thermal fluctuations? *Astrophys. J.* **698**, 895–910 (2009).
2. Pojmanski, G. The All Sky Automated Survey. Catalog of Variable Stars. I. 0 h – 6 h Quarter of the Southern Hemisphere. *Acta Astronom.* **52**, 397–427 (2002).
3. Udalski, A. *et al.* The optical gravitational lensing experiment. Discovery of the first candidate microlensing event in the direction of the Galactic Bulge. *Acta Astronom.* **43**, 289–294 (1993).
4. Rau, A. *et al.* Exploring the optical transient sky with the Palomar Transient Factory. *Publ. Astron. Soc. Pacif.* **121**, 1334–1351 (2009).
5. Sesar, B. *et al.* Exploring the variable sky with LINEAR. I. Photometric recalibration with the Sloan Digital Sky Survey. *Astron. J.* **142**, 190–202 (2011).
6. Valtonen, M. J. *et al.* A massive binary black-hole system in OJ287 and a test of general relativity. *Nature* **452**, 851–853 (2008).
7. Valtonen, M. J., Lehto, H. J., Takalo, L. O. & Sillanpää, A. Testing the 1995 binary black hole model of OJ287. *Astrophys. J.* **729**, 33–38 (2011).
8. Ju, W., Greene, J. E., Rafikov, R. R., Bickerton, S. J. & Badenes, C. Search for supermassive black hole binaries in the Sloan Digital Sky Survey spectroscopic sample. *Astrophys. J.* **777**, 44–59 (2013).
9. Shen, Y., Liu, X., Loeb, A. & Tremaine, S. Constraining sub-parsec binary supermassive black holes in quasars with multi-epoch spectroscopy. I. The general quasar population. *Astrophys. J.* **775**, 49–71 (2013).
10. Tsalmantza, P., Decarli, R., Dotti, M. & Hogg, D. W. A systematic search for massive black hole binaries in the Sloan Digital Sky Survey spectroscopic sample. *Astrophys. J.* **738**, 20–28 (2011).
11. Drake, A. J. *et al.* First results from the Catalina Real-Time Transient Survey. *Astrophys. J.* **696**, 870–884 (2009).
12. Djorgovski, S. G. *et al.* in *The First Year of MAXI: Monitoring Variable X-ray Sources* (eds Mihara, T. & Kawai, N.) 263–268 (JAXA Special Publication, 2010); available at http://maxi.riken.jp/FirstYear/proceedings/procindex.html.
13. Mahabal, A. A. *et al.* Discovery, classification, and scientific exploration of transient events from the Catalina Real-time Transient Survey. *Bull. Astron. Soc. India* **39**, 387–408 (2011).
14. Hinshaw, G. *et al.* Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results. *Astrophys. J.* **208** (suppl.), 19–43 (2013).
15. Garcia, A., Sodré, L., Jablonski, F. J. & Terlevich, R. J. Optical monitoring of quasars — I. Variability. *Mon. Not. R. Astron. Soc.* **309**, 803–816 (1999).
16. Eggers, D., Shaffer, D. B. & Weistrop, D. Optical variability of radio-luminous PG quasars. *Astron. J.* **119**, 460–468 (2000).
17. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. Hubble Space Telescope images of nearby luminous quasars. II. Results for eight quasars and tests of the detection sensitivity. *Astrophys. J.* **450**, 486–500 (1995).
18. McLure, R. J. *et al.* A comparative HST imaging study of the host galaxies of radio-quiet quasars, radio-loud quasars and radio galaxies — I. *Mon. Not. R. Astron. Soc.* **308**, 377–404 (1999).
19. Hutchings, J. B., Morris, S. C., Gower, A. C. & Lister, M. L. Correlated optical and radio structure in the QSO 1302–102. *Publ. Astron. Soc. Pacif.* **106**, 642–645 (1994).
20. Benítez, E. *et al.* The close environment of OJ 287: underlying nebulosity and a possible optical jet? *Astrophys. J.* **464**, L47–L50 (1996).
21. Jackson, N. *et al.* Monitoring of active galactic nuclei. I — The quasars 1302-102 and 1217+023. *Astron. Astrophys.* **262**, 17–25 (1992).
22. Shen, Y. & Loeb, A. Identifying supermassive black hole binaries with broad emission line diagnosis. *Astrophys. J.* **725**, 249–260 (2010).
23. Gaskell, C. M. Close supermassive binary black holes. *Nature* **463**, E1 (2010).
24. Lu, J.-F. & Zhou, B.-Y. Observational evidence of jet precession in galactic nuclei caused by accretion disks. *Astrophys. J.* **635**, L17–L20 (2005).
25. Shen, Y. *et al.* A catalog of quasar properties from Sloan Digital Sky Survey data release 7. *Astrophys. J.* **194** (suppl.), 45–65 (2011).
26. Farris, B. D., Duffell, P., MacFadyen, A. I. & Haiman, Z. Binary black hole accretion from a circumbinary disk: gas dynamics inside the central cavity. *Astrophys. J.* **783**, 134–145 (2014).
27. Rafikov, R. R. Structure and evolution of circumbinary disks around supermassive black hole binaries. *Astrophys. J.* **774**, 144 (2013).
28. Tremaine, S. & Davis, S. W. Dynamics of warped accretion discs. *Mon. Not. R. Astron. Soc.* **441**, 1408–1434 (2014).
29. Bogdanović, T., Smith, B. D., Sigurdsson, S. & Eracleous, M. Modeling of emission signatures of massive black hole binaries. I. Methods. *Astrophys. J.* **174** (suppl.), 455–480 (2008).

Author Contributions M.J.G. performed the analysis and wrote the paper. S.G.D. is the PI of the CRTS survey and obtained the Keck spectrum. E.G. obtained and reduced the near-infrared data and provided the Balmer and Paschen line fits. D.S. reduced the Keck data. A.J.D. is the co-PI of the CRTS survey and provided the CRTS data. S.L. and E.C. are the PIs of the CSS survey. All authors contributed to the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.G. (mjg@caltech.edu).

METHODS

CRTS data. The Catalina Real-time Transient Survey (CRTS) makes use of the Catalina Sky Survey which began in 2004, operated by the Lunar and Planetary Laboratory at the University of Arizona, and uses three telescopes (designated CSS, MLS and SSS) to cover the sky between declination $\delta = -75°$ and $+65°$ ($\sim80\%$ of the sky) in order to discover near-Earth objects and potentially hazardous asteroids. The full Catalina surveys data set contains time series for approximately 500 million sources to a limiting magnitude of $V \approx 20$ with an average of ~250 observations over a baseline of 9 years per source. CRTS operates an open data policy, and the data are publicly available at: http://catalinadata.org.

For photometric calibration we combine observations taken with all three Catalina Sky Survey telescopes since no difference was found between these systems. This is not surprising since each telescope specifically uses the same type of 4k × 4k CCD camera and observations with all three telescopes are calibrated using the same software pipeline. All Catalina observations are transformed to Johnson V based on 50–100 stars selected as G-type stars using 2MASS[30] colours. For bright stars, this photometry provides repeated photometry accurate to ~0.05[31]. However, as the photometry is unfiltered, there are significant variations with object colour.

The 2007 Landolt *UBVRI* standard star catalogue[32] provides 109 stars centred near declination $-50°$ in the magnitude range $10.4 < V < 15.5$ and in the colour index range $-0.33 < (B - V) < 1.66$, while the 2009 Landolt catalogue[33] provides 202 standard stars along the celestial equator in the magnitude range $8.90 < V < 16.30$, and the colour index range $-0.35 < (B - V) < 2.30$, along with 393 standard stars from previous standard star catalogues. In total there are 445 Catalina light curves matching Landolt standards. On average each standard is measured 134 times. A handful of stars that appeared to exhibit significant variability were removed.

Median magnitudes were calculated for each light curve and these were then used to determine the following transformation equations between Johnson V and Catalina V_{CSS}:

$$V = V_{CSS} + 0.31 \times (B - V)^2 + 0.04$$

$$V = V_{CSS} + 0.91 \times (V - R)^2 + 0.04$$

$$V = V_{CSS} + 1.07 \times (V - I)^2 + 0.04$$

The dispersion in the fits to these transformations are 0.059, 0.056 and 0.063 mag, respectively, for $V < 16$.

Candidate selection. We applied the weighted wavelet transform[34] and the z-transform discrete correlation function (ZDCF)[35] to the CRTS light curves of spectroscopically confirmed quasars. Both of these algorithms can detect (quasi-) periodic behaviour in irregularly sampled data. We define the period of the quasar from the largest peak in the ZDCF between the second and third zero-crossings of a Gaussian process model fit to the ZDCF[36]. We have verified that this agrees with periods determined by other methods. The period uncertainty is defined as:

$$\sigma_P = \frac{1.483 \times MAD}{\sqrt{N - 1}}$$

where MAD is the median of the absolute deviations from the median of the time intervals between successive peaks in the ZDCF and N is the total number of peaks considered.

We only considered those objects whose wavelet peak significance places them in the top quartile of the data set (in terms of significance). Most kinematically-caused variations should manifest as a (near-)Keplerian signal and so we also use the r.m.s. scatter around an expected sinusoidal waveform (best-fit truncated Fourier series of up to 6 terms) with the ZDCF period to identify those objects most closely exhibiting the expected behaviour. We excluded those quasars where the scatter is greater than the 1σ lower limit on the median absolute deviation of the light curve, that is, r.m.s./MAD > 0.67, since we need to account for the intrinsic variability of the quasar as well. We also restricted our selection to candidates with temporal coverage of more than 1.5 cycles, assuming the ZDCF period.

Simulated data. For each quasar in our data set, we have generated a simulated light curve assuming that it follows a damped random walk (DRW) model. Using the actual observation times, t_i, we replace the observed magnitudes with those that would be expected under a DRW model. The magnitude $X(t)$ at a given timestep Δt from a previous value $X(t - \Delta t)$ is drawn from a Gaussian distribution with mean and variance[1]:

$$E(X(t)|X(t - \Delta t)) = e^{-\Delta t/\tau} X(t - \Delta t) + b\tau(1 - e^{-\Delta t/\tau})$$

$$\mathrm{Var}(X(t)|X(t - \Delta t)) = \frac{\tau\sigma^2}{2}\left[1 - e^{-2\Delta t/\tau}\right]$$

We add a Gaussian deviate normalized by the photometric error associated with the magnitude to be replaced at each time t to incorporate measurement uncertainties into the mock light curves. For each light curve, we set $b\tau$ to its median value and use the rest frame DRW fitting functions[37]:

$$\log f = A + B \log\left(\frac{\lambda_{RF}}{4000\text{Å}}\right) + C(M + 23) + D \log\left(\frac{M_{BH}}{10^9 M_\odot}\right)$$

where $(A, B, C, D) = (-0.51, -0.479, 0.113, 0.18)$ for $f = SF_\infty = \tau\sigma^2/2$ and $(A, B, C, D) = (2.4, 0.17, 0.03, 0.21)$ for $f = \tau$. M is the absolute magnitude of the quasar and λ_{RF} is the rest-frame wavelength of the filter. The mass of the black hole is either the measured virial mass[38] or is drawn from a Gaussian distribution[37]:

$$p(\log M_{BH}|M) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\log M_{BH} - \mu)^2}{2\sigma^2}\right]$$

where $\mu = 2.0 - 0.27M$ and $\sigma = 0.58 + 0.011M$.

Statistical significance. To assess the statistical significance of the detection of PG 1302−102, we generated 1,000 simulated light curves as above. From these we determined the mean weighted wavelet power spectrum as a function of time and frequency and its variance. The dominant signal in the observed WWZ spectrum of PG 1302−102 is then seen to be ~40 times above the corresponding mean DRW value in terms of the expected standard deviation.

We have also performed a periodicity analysis of the light curve of PG 1302−102 using the generalized Lomb-Scargle method which shows a statistically significant peak at the same period identified by the wavelet and autocorrelation analyses. The false alarm probability is $\ll 10^{-13}$ — the 10^{-13} level is $P(\omega) = 0.335$ and the observed peak is at $P(\omega) = 0.818$.

Theoretical predictions. Simple disk models for circumbinary gas and the binary–disk interaction have been used[39] to consider the number of SMBH binaries expected in a variety of surveys, assuming that such objects are in the final gravitational-wave-dominated phase of coalescence (this equates to separations less than ~0.01 pc for a $10^8 M_\odot$ SMBH binary). This approach has been combined[40] with merger tree assembly models to similarly predict the number of expected SMBH binaries at wider separations where spectral line shifts may be seen (this equates to separations greater than ~0.2 pc for a $10^8 M_\odot$ SMBH binary). The latter shows that in a sample of 10,000 quasars at $z < 0.7$, there should be ~10 objects and this number increases by a factor of ~5–10 for $z < 1$. We note, however, that these theoretical arguments are still subject to considerable uncertainties; for example, if the final parsec problem cannot be resolved then there will not be any binaries in the ~0.01 pc regime.

Assuming a limiting magnitude of $V \approx 20$, a detectable range of orbital periods from 20–300 weeks (spanning both GW- and gas-dominated regimes), a survey sky coverage of 2π steradians, and a redshift range of 0.5–4.5, we would expect 450 SMBH binaries following these approaches. Our finding of 20 candidates from a sample of 240,000 quasars is therefore conservative. 89,000 quasars in our sample also have virial black-hole mass estimates[38] (23% at $z > 2$) and if we assumed that each of these was a SMBH binary with a separation of 0.01 pc then the CRTS temporal baseline is sufficient to detect 1.5 cycles or more in 63% of them (including 55% of the $z > 2$ population). Our search is therefore sensitive to a large fraction of the close SMBH binary population.

We note that our approach assumes that periodicity associated with SMBH binaries manifests in a Keplerian form. If there is a larger set of non-Keplerian periodic SMBH binaries, either flaring, such as OJ 287, or not, then the 20 objects we have identified may be a small sample of the total close binary SMBH population.

Archival data. *LINEAR data*[5]. These were calibrated with pre-release photometry from Pan-STARRS using the g, r and i bandpasses. Comparison stars with instrumental magnitudes (ccd_mag) between 14 and 17 were selected within 0.1° of PG 1302−102 in LINEAR images. $g - i$ colours were used to compute an r-band correction so that a calibration star with $g - i = 0$ has an instrumental magnitude, ccd_mag = r. Zeropoints for each frame were then derived based on these stars. The reported bandpass for the calibrated magnitudes is therefore approximately r. Magnitude errors are computed by SEXTRACTOR, with typical r.m.s. errors between frames of ~0.1.

ASAS data. The nominal limiting magnitude for ASAS[2] is $I \approx 13$ and so PG 1302−102 is very close to the detection threshold. The low signal-to-noise ratio for such an object is the primary cause of the large degree of scatter seen in ASAS data for this source.

Historic data. Such data for PG 1302 from previous quasar monitoring campaigns is available in the literature[15,16]. To put all data on the same photometric scale, offsets were applied to account for differences in the photometric systems used. Region of temporal overlap between a pair of data sets were used to derive offsets so that both data had the same median value. Where no temporal overlap exists, the phased light curve was used to determine the median offset.

Earlier individual photometric observations also exist of PG 1302−102 but the observational errors on these are typically ~0.1 mag and so it is difficult to determine

whether they agree with the extrapolated behaviour. They also tend to be in different passbands which requires colour terms to convert to the V-passband to which CRTS is calibrated. However, colour terms for quasars are known to vary ("bluer when brighter") so a constant value cannot be assumed (the quoted $(B - V)$ values for PG 1302-02 have a range of at least 0.2 mag), which introduces an additional error to the transformed magnitude. Such historical data are thus of limited utility.

Spectroscopic data. An optical spectrum was obtained using the Double Spectrograph on the Hale 200-inch telescope at Palomar Observatory on UT 2014 April 22. We obtained two 250 s exposures in cloudy conditions using the 1.0″ wide slit, the 5,500 Å dichroic, the 600 λ mm^{-1} grating on the blue arm ($\lambda_{\rm blaze} = 4,000$ Å), and the 316 λ mm^{-1} grating on the red arm ($\lambda_{\rm blaze} = 7,500$ Å).

On UT 2014 May 26, we obtained additional spectroscopy of PG 1302−102 using the Low Resolution Imaging Spectrometer (LRIS)[41] on the Keck I telescope. We obtained two 300-s exposures in non-photometric conditions, using the 1.5″ slit, the 5,600 Å dichroic, the 600 λ mm^{-1} grism on the blue arm ($\lambda_{\rm blaze} = 4,000$ Å) and the 400 λ mm^{-1} grating on the red arm ($\lambda_{\rm blaze} = 8,500$ Å).

These set-ups provided moderate resolution spectra across the entire optical window, 3,100 Å to 1 μm. The data from both telescopes were reduced using standard procedures and calibrated using observations of standard stars obtained on the same (non-photometric) nights.

Near-infrared. We obtained a near-infrared spectrum of PKS 1302−102 with the TripleSpec instrument on the Hale 200-inch telescope at Palomar Observatory on UT 2014 April 15. Conditions were clear and the seeing was ∼1 arcsec. The source was observed at an airmass of 1.3890 and was observed for four 300-s exposures in an ABBA dither pattern for a total of 20 min of on-source exposure. A spectrum of telluric standard A0V star HD 112304 was obtained immediately following the source spectrum at an airmass difference of 0.15. We reduced the data using a modified version of the Spextool data reduction package[42] which also performs a telluric correction using the standard star spectrum[43].

Figure 3 shows the combined optical and near-infrared spectrum, in which we scaled the (non-photometric) optical spectrum by a factor of 2.86 to meet the near-infrared spectrum in their mutually overlapping wavelength region (0.96–1.05 μm). The change in flux is probably primarily due to a combination of weather variations and slit losses. The Balmer lines, Hβ and Hα, and Paschen lines, Paβ and Paα, are marked with vertical dashed lines.

System parameters. To estimate the (total) black-hole mass for this source, we used the standard method of single epoch virial black-hole estimation[44] and adapted relations derived for Paschen lines in the near-infrared[45]. To determine the full-width at half-maximum (FWHM) of the broad line component of the four lines marked in Fig. 4, we applied a multi-component line fitting technique[46,47] in which we model the narrow-line component of the line profiles by first fitting the [O III] 4,959, 5,007 Å lines and fixing the width of a narrow line component in each profile. For Hα we also include the [N II] 6,548, 6,583 Å doublet, fixed at a flux ratio of 2.96. The broad component can be modelled by up to three Gaussians. If the ratio of χ^2 values for successive fits is greater than 0.8, an additional component is added. The near-infrared spectrum has an error array which is used in the line modelling and parameter estimation (see below). For the optical spectrum, an error array is estimated from the median difference between adjacent pixels (R. White, personal communication).

The measured values from the 2014 data are: $4,450 \pm 150$ km s^{-1} for Hβ, $2,520 \pm 30$ km s^{-1} for Paα, and $3,200 \pm 20$ km s^{-1} for Paβ. Errors on the FWHM were computed using a Monte Carlo approach: the best-fit model for a line was perturbed with a random draw from the error array at each wavelength element and a new fit made. This was repeated 100 times for each line and the standard deviation of the broad line component FWHM used as the error. These values give a (total) mass for the SMBH of $\log(M/M_\odot) = 8.8 \pm 0.6$ (Hβ), 8.5 ± 0.1 (Paβ), and 8.4 ± 0.1 (Paα). Previously reported values of the FWHM of Hβ give a range of estimates for the SMBH mass in the literature using various techniques of $\log(M/M_\odot) = 8.3$–9.4, so our results are consistent with these.

We note that the spectral fits of Hβ and Paβ are not perfect. However, the uncertainties in the FWHM and continuum luminosity that this introduces are small compared to the broad range of SMBH mass estimated from the different lines.

Alternative interpretations. We present further discussion here on the alternative interpretations considered. We note that none of the objects mentioned pass our candidate selection criteria.

Jet related. The optical flux could be the superposition of thermal contribution from the accretion disk with non-thermal contribution produced by the underlying jet. The flux density of the jet (in the optically thin regime) will be boosted in the observer's frame relative to the co-moving frame:

$$S_j(v) = S_j'(v)\delta(\phi,\gamma)^{p+\alpha}$$

where α is the spectral index ($S_j(v) \propto v^{-\alpha}$) and $p = 2$ for a continuous jet. The jet is precessing with constant angular velocity ω, has an opening angle Ω and an axis

defined by the angles ϕ_0 (between the jet axis and the line of sight) and η_0 (the position angle in the plane of the sky):

$$\sin^2 \phi = (\sin \Omega \cos \omega t + \cos \Omega \sin \phi_0 \sin \eta_0)^2$$
$$+ (\sin \Omega \cos \phi_0 \sin \omega t + \cos \Omega \sin \phi_0 \cos \eta_0)^2$$

Assuming a constant Lorentz factor γ for the relativistic bulk motion of the jet, $\gamma = (1 - \beta^2)^{-1/2}$, the Doppler factor is given by: $\delta = \gamma^{-1}(1 - \beta\cos\phi)^{-1}$. Modelling the light curve in this way, we get best-fit parameters of: $\gamma = 5.4 \pm 0.1$, $\Omega = 0.5°$ ± 0.1, $\phi_0 = 5.0° \pm 0.2$, and $\eta_0 = 0.6° \pm 1.4$ (assuming $\alpha = 1.66$).

A number of radio-loud quasars have been reported[48–50] as showing periodic variability in their radio light curves. While a SMBH binary could explain this, a more likely explanation is shock interaction with a helical jet or precession of a jet. However, the optical light curves of these objects (see Extended Data Fig. 1) do not show the distinctive behaviour seen in that of PG 1302−102 suggesting that a different physical mechanism is more likely. We note as well that of the 20 objects in our full sample showing optical periodicity, only 3 are associated with a radio source.

Warped accretion disks. These have been observed in a handful of AGN[51–54] and the suggestion here is that as a warp precesses, it could obscure a small amount of continuum emission which would then appear quite regular. Again there is no indication of any periodic behaviour in the CRTS light curves available for known objects with warped disks (see Extended Data Fig. 2) similar to that seen in PG 1302−102.

PG 1302−102 shows a 14% variation in flux, which would suggest that the size of the warp in the disk is quite large. This would also be an orientation-dependent phenomenon and as the source is a blazar, its accretion disk should be oriented close to face-on to us and so any obscuring factor should be limited in effect. We also note that many stellar systems with warped accretion disks are resolvable binary systems.

30. Skrutskie, M. F. *et al.* The Two Micron All Sky Survey. 2MASS. *Astron. J.* **131,** 1163–1183 (2006).
31. Larson, S. *et al.* The CSS and SSS NEO surveys. *Bull. Am. Astron. Soc.* **35,** 982 (2003).
32. Landolt, A. U. UBVRI photometric standard stars around the sky at −50° declination. *Astron. J.* **133,** 2502–2523 (2007).
33. Landolt, A. U. UBVRI photometric standard stars around the celestial equator: updates and additions. *Astron. J.* **137,** 4186–4269 (2009).
34. Foster, G. Wavelets for period analysis of unevenly sampled time series. *Astron. J.* **112,** 1709–1729 (1996).
35. Alexander, T. Improved AGN light curve analysis with the z-transformed discrete correlation function. Preprint at http://arXiv.org/abs/1302.1508 (2013).
36. McQuillan, A., Aigrain, S. & Mazeh, T. Measuring the rotation period distribution of field M dwarfs with Kepler. *Mon. Not. R. Astron. Soc.* **432,** 1203–1216 (2013).
37. MacLeod, C. L. *et al.* Modeling the time variability of SDSS Stripe 82 quasars as a damped random walk. *Astrophys. J.* **721,** 1014–1033 (2010).
38. Shen, Y. *et al.* A catalog of quasar properties from Sloan Digital Sky Survey data release 7. *Astrophys. J.* **194** (Suppl.), 45–65 (2011).
39. Haiman, Z., Kocsis, B. & Menou, K. The population of viscosity- and gravitational wave-driven supermassive black hole binaries among luminous active galactic nuclei. *Astrophys. J.* **700,** 1952–1969 (2009).
40. Volonteri, M., Miller, J. M. & Dotti, M. Sub-parsec supermassive binary quasars: expectations at z<1. *Astrophys. J.* **703,** L86–L89 (2009).
41. Oke, J. B. *et al.* The Keck Low-Resolution Imaging Spectrometer. *Publ. Astron. Soc. Pacif.* **107,** 375–385 (1995).
42. Cushing, M. C., Vacca, W. D. & Rayner, J. T. Spextool: a spectral extraction package for SpeX, a 0.8-5.5 micron cross-dispersed spectrograph. *Publ. Astron. Soc. Pacif.* **116,** 362–376 (2004).
43. Vacca, W. D., Cushing, M. C. & Rayner, J. T. A method of correcting near-infrared spectra for telluric absorption. *Publ. Astron. Soc. Pacif.* **115,** 389–409 (2003).
44. Shen, Y. The mass of quasars. *Bull. Astron. Soc. India* **41,** 61–115 (2013).
45. Kim, D., Im, M. & Kim, M. New estimators of black hole mass in active galactic nuclei with hydrogen Paschen lines. *Astrophys. J.* **724,** 386–399 (2010).
46. Greene, J. E. & Ho, L. C. Active galactic nuclei with candidate intermediate-mass black holes. *Astrophys. J.* **610,** 722–736 (2004).
47. Glikman, E. *et al.* The FIRST-2MASS red quasar survey. *Astrophys. J.* **667,** 673–703 (2007).
48. Raiteri, C. M. *et al.* Optical and radio variability of the BL Lacertae object AO 0235+16: a possible 5-6 year periodicity. *Astron. Astrophys.* **377,** 396–412 (2001).
49. Fan, J. H. *et al.* Optical periodicity analysis for radio selected BL Lacertae objects (RBLs). *Astron. Astrophys.* **381,** 1–5 (2002).
50. Kudryavtseva, N. A. *et al.* A possible jet precession in the periodic quasar B0605-085. *Astron. Astrophys.* **526,** A51–A64 (2011).
51. Greenhill, L. J. *et al.* A warped accretion disk and wide-angle outflow in the inner parsec of the Circinus Galaxy. *Astrophys. J.* **590,** 162–173 (2003).
52. Herrnstein, J. R., Moran, J. M., Greenhill, L. J. & Trotter, A. S. The geometry of and mass accretion rate through the maser accretion disk in NGC 4258. *Astrophys. J.* **629,** 719–738 (2005).
53. Kondratko, P. T., Greenhill, L. J. & Moran, J. M. The parsec-scale accretion disk in NGC 3393. *Astrophys. J.* **678,** 87–95 (2008).
54. Kuo, C. Y. *et al.* The Megamaser Cosmology Project. III. Accurate masses of seven supermassive black holes in active galaxies with circumnuclear megamaser disks. *Astrophys. J.* **727,** 20–34 (2011).
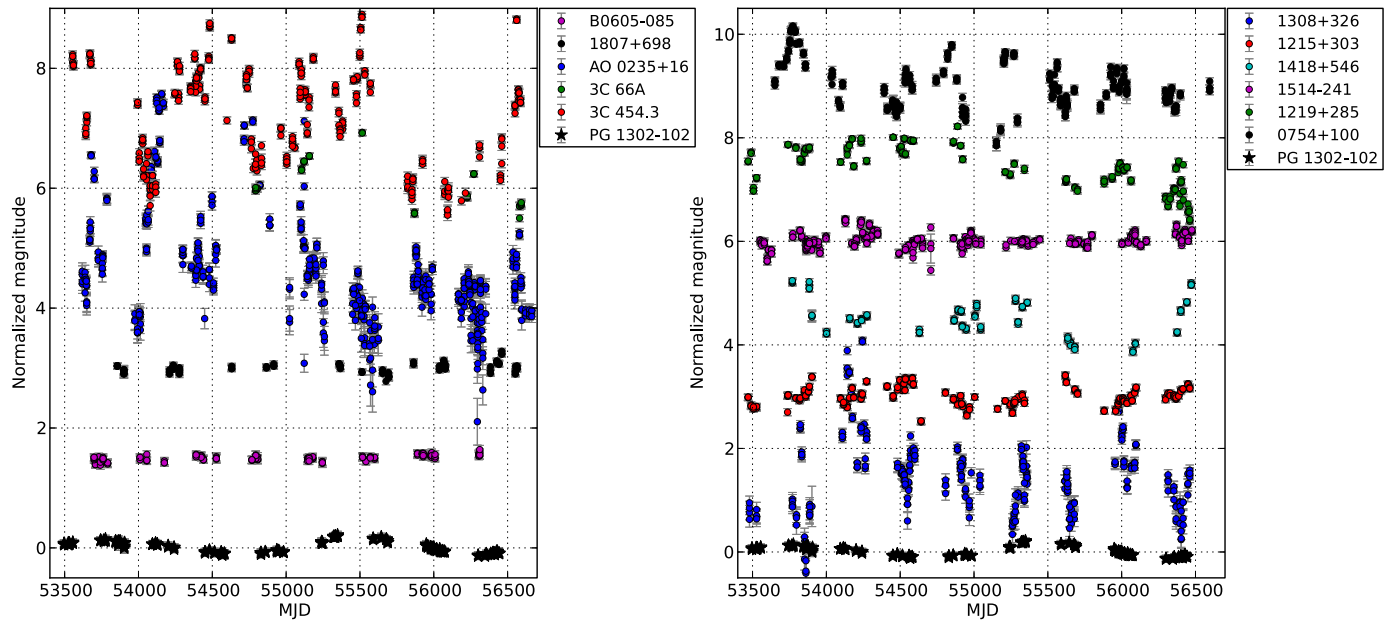
Extended Data Figure 1 | The optical light curves of quasars showing radio periodicity. Shown are the CRTS light curves for 11 quasars reported[49,50] to show periodicity in their radio emission. Each light curve has been normalized to zero mean and individual curves are offset by a constant of 1.5 mag from each other. The data are split across two panels for ease of viewing. Error bars shown are standard 1σ photometric errors. The CRTS light curve of PG 1302−102 (solid black stars) is also shown for comparison.
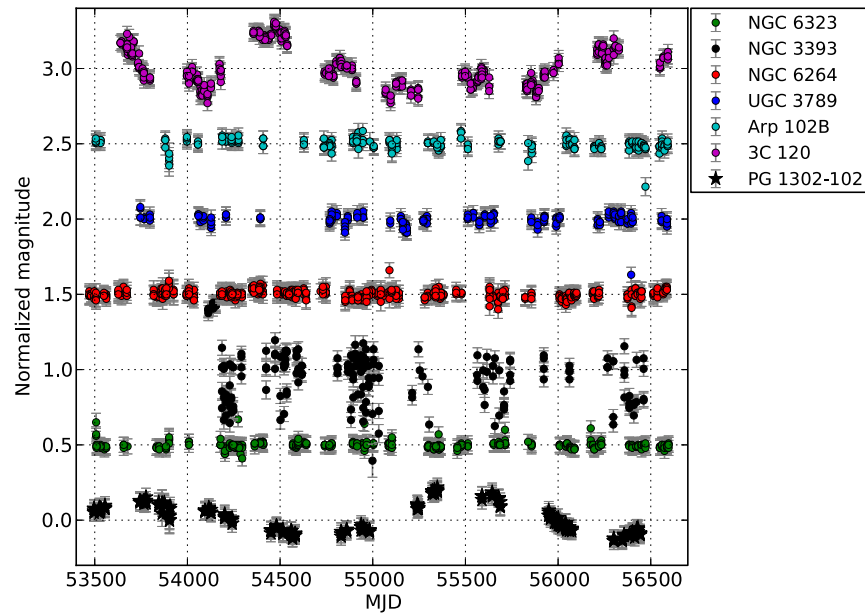
Extended Data Figure 2 | The optical light curves of quasars with warped accretion disks. Shown are the CRTS light curves for 6 quasars reported[52–55] to have warped accretion disks. Each light curve has been normalized to zero mean and individual curves are offset by a constant of 0.5 mag from each other. Error bars shown are standard 1σ photometric errors. The CRTS light curve of PG1302−102 (solid black stars) is also shown for comparison.

Brittle intermetallic compound makes ultrastrong low–density steel with large ductility

Sang–Heon Kim[1], Hansoo Kim[1] & Nack J. Kim[1]

Although steel has been the workhorse of the automotive industry since the 1920s, the share by weight of steel and iron in an average light vehicle is now gradually decreasing, from 68.1 per cent in 1995 to 60.1 per cent in 2011 (refs 1, 2). This has been driven by the low strength-to-weight ratio (specific strength) of iron and steel, and the desire to improve such mechanical properties with other materials. Recently, high-aluminium low-density steels have been actively studied as a means of increasing the specific strength of an alloy by reducing its density[3–5]. But with increasing aluminium content a problem is encountered: brittle intermetallic compounds can form in the resulting alloys, leading to poor ductility. Here we show that an FeAl-type brittle but hard intermetallic compound (B2) can be effectively used as a strengthening second phase in high-aluminium low-density steel, while alleviating its harmful effect on ductility by controlling its morphology and dispersion. The specific tensile strength and ductility of the developed steel improve on those of the lightest and strongest metallic materials known, titanium alloys. We found that alloying of nickel catalyses the precipitation of nanometre-sized B2 particles in the face-centred cubic matrix of high-aluminium low-density steel during heat treatment of cold-rolled sheet steel. Our results demonstrate how intermetallic compounds can be harnessed in the alloy design of lightweight steels for structural applications and others.

There is increasing demand for a broad range of structural materials for environmentally benign, energy-efficient, lightweight engineering systems. The balance of lightness, strength and ductility in metallic alloys has been explored since the Bronze Age. Unfortunately, strength and ductility are mutually exclusive. Further, higher specific strength is extremely difficult to achieve because strength invariably decreases as density is reduced. However, the alloying of iron with aluminium results in higher specific strength even though density is reduced. Until now, low-density steel has been studied[3–7] mostly in systems based on Fe–Al and Fe–Al–Mn–C. In particular, low-density steel with fairly high specific strength has been produced with alloys based on Fe–Al–Mn–C (the so-called TRIPLEX steels) using a microstructure consisting of austenite (face-centred cubic) matrix and finely dispersed nanometre-sized κ-carbides of the (Fe,Mn)$_3$AlC type[3,6,7]. However, the level of specific strength attainable by this microstructure was not comparable to those of light materials such as aluminium and titanium alloys. This was due to the low strain hardening rate of the Fe–Al–Mn–C alloys containing κ-carbides, which are easily shearable by gliding dislocations[3].

One of the general concepts employed until now in the alloy design of Fe–Al–Mn–C-based, high-aluminium, low-density steel has been the suppression of 'brittle' intermetallic compound formation by stabilizing the 'ductile' austenite matrix[3,6–8] (this stabilization is achieved by alloying carbon and manganese). Instead, here we have actively utilized the brittle intermetallic compound B2 by modifying its morphology in the steel matrix. Despite their poor plasticity at ambient temperature in the bulk state, FeAl-based intermetallic compounds offer an attractive combination of physical and mechanical properties such as low density and good corrosion, oxidation and/or wear resistance[9–13]. To take advantage of B2, we devised an alloy design in which B2 is dispersed as a second phase in the austenite matrix on the basis of the 'divide and rule'

principle, which is analogous to harnessing 'brittle' martensite as a strengthening second phase in the ferrite (body-centred cubic) matrix of dual-phase steels[14,15].

A common method of uniformly distributing fine particles in a matrix is to make the best use of highly potent nucleation sites for inducing the precipitation of the particles. In this study, potential nucleation sites for B2 during annealing of wrought sheet steel include (1) grain boundaries or edges of recrystallized austenite crystals and (2) deformation shear bands, which are common in hot- or cold-worked low-density steel[3,5,7,16,17]. To expand the stability domain of B2 above the recrystallization temperature (normally, 800–900 °C) of deformed austenite, the alloying recipe of an austenitic low-density steel was modified by adding 5 weight per cent nickel (Ni), which is one of the most effective elements for forming B2 with aluminium[18,19]. The addition of Ni to low-density steel may appear to conflict with the collective wisdom of ferrous alloy design; Ni has been regarded merely as a well-known austenite stabilizer like Mn and C; and Ni has been little noticed in low-density steel design, mainly because it is not a critical determinant of the density in ferrous alloys.

The steel under investigation here was produced using an induction melting furnace. About 40 kg was melted in a protective argon atmosphere and cast to a rectangular ingot (300 mm width, 80 mm thickness, 240 mm length). After homogenization treatment at 1,150 °C for 2 h, the ingot was hot-rolled with a starting temperature of 1,050 °C to hot strips 3 mm in thickness. Then, the hot-rolled strips were cold-rolled to final sheets 1 mm in thickness. The cold-rolled sheets were annealed at 870–900 °C for 2–60 min and immediately water-quenched or continuously cooled down to 25 °C at the rate of 30 °C s^{-1}. The chemical composition of the present steel and selected reference materials was measured by wet chemical analysis and is given in Extended Data Table 1.

The reference materials—the press hardening steel and the titanium alloy of Ti6Al4V—were obtained from materials suppliers. Flat tensile specimens with a gauge dimension of 12.5 mm width by 50 mm length by 1 mm thickness were machined in such a way that the tensile axis is parallel to the rolling direction. Tensile tests were carried out at an initial strain rate of 10^{-3} s^{-1} at ambient temperature. The density of our steel and of the selected reference materials was measured by pycnometry. The thermal treatment conditions and the properties of our steel and of the reference materials are listed in Extended Data Table 2. The phases present in the microstructure of our steel were identified by X-ray diffraction (Extended Data Fig. 1). Microstructural investigations were performed by scanning electron microscopy (SEM) and transmission electron microscopy (TEM) coupled with energy dispersive spectroscopy (EDS).

Figures 1a and b show SEM images of the novel high-specific-strength steel (HSSS) of the present study, comparing the microstructures of as-cold-rolled and annealed sheets, respectively (see Extended Data Tables 1 and 2 and Extended Data Fig. 1). The as-cold-rolled microstructure consists of austenite matrix and B2 stringer bands parallel to the rolling direction (Fig. 1a), but the microstructure dramatically changes after 15 min of annealing at 900 °C, precipitating out fine B2 particles in between the B2-stringer bands in the steel matrix (Fig. 1b). B2 in the

[1]Graduate Institute of Ferrous Technology, POSTECH, Pohang 790-784, South Korea.
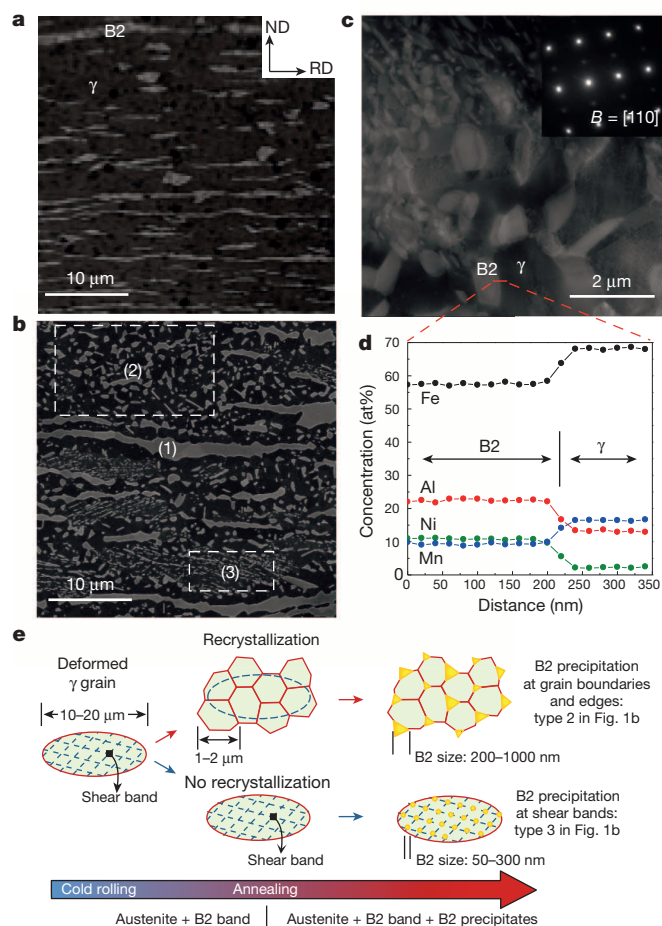
Figure 1 | Precipitation of B2 particles during annealing of cold rolled Fe–10%Al–15%Mn–0.8%C–5%Ni (weight per cent) high-specific-strength steel. **a**, As-cold-rolled microstructure consisting of austenite matrix (γ) and B2 stringer bands. RD, rolling direction; ND, normal direction. **b**, Annealed microstructure having fine B2 precipitates in between the retained B2 bands in austenite matrix. **c**, Scanning TEM image of the annealed high-specific-strength steel (HSSS) showing morphologies of B2 particles. The inset shows the selected area diffraction pattern of a B2 precipitate. **d**, Partitioning of alloying elements between B2 precipitate and austenite matrix. **e**, Sketches illustrating the formation mechanism of B2 precipitates of types 2 and 3 in **b**.

annealed sheet (Fig. 1b) has three different morphologies, comprising retained stringer bands (type 1), fine particles of size 200–1,000 nm (type 2) and finer particles of size 50–300 nm (type 3). A scanning TEM image of Fig. 1c clearly shows the size difference of B2 particles of types 2 and 3. The inset of Fig. 1c shows the selected area diffraction pattern of a B2 particle. Partitioning of each element between B2 and austenite is illustrated in Fig. 1d, showing EDS composition profiles of Fe, Al, Ni and Mn across the interface of B2 and austenite (γ). Al and Ni are enriched in B2 while the Mn content is higher in austenite (γ). Figure 1e schematically illustrates the formation mechanism of B2 precipitates of types 2 and 3. If a cold-worked metal is heated to a sufficiently high temperature, strain-free new grains are formed by recrystallization, replacing deformed ones[20]. Owing to the inhomogeneous nature of plastic deformation at a fine scale[20,21], the recrystallization kinetics of microscopically local areas can be substantially different in deformed steel matrix. When the local recrystallization proceeds quickly during annealing, B2 precipitates at recrystallized fine austenite grain boundaries (or edges) to form the type 2 microstructure. Otherwise, B2 precipitates along the shear bands in non-recrystallized coarse austenite grains, resulting in the type 3 microstructure.

Compared in Fig. 2 are the room-temperature tensile properties of the ductile HSSS of the present study and selected metallic alloys of high specific strength[6,7,22–27] (Extended Data Tables 1 and 2). Figure 2a shows the representative stress–strain curves of HSSS annealed under various conditions in comparison with Fe-Al-Mn-C-based high-aluminium, low-density steel[6,7], a commercial titanium alloy (Ti6Al4V) and a press hardening (fully martensitic) boron steel[28]. Flow curves of HSSS show large ductility and phenomenally high strain hardening capability even at ultrahigh yield strength levels of over 1 GPa. Figure 2b shows the relation of specific ultimate tensile strength (SUTS) versus elongation to fracture. The HSSS shows an exceptional combination of specific strength (SUTS) and elongation compared with other high-specific-strength alloys. Figure 2c shows the relation of specific yield strength (SYS) versus the increase in density-compensated tensile toughness during uniform elongation (UE). The expression UE(SUTS − SYS) corresponds to the increase of specific toughness due to strain hardening during uniform plastic deformation, that is, for the uniform part of the elongation before necking begins after yielding. Owing to the high strain hardening capability, HSSS shows extraordinarily high specific tensile toughness even at ultrahigh levels of SYS over 150 MPa g^{-1} cm^3.

Scanning TEM images in Fig. 3 show the interaction of dislocations with B2 particles in HSSS that had previously been deformed by 0.5%. Dislocations are piled up (Fig. 3a) or bowing out (Fig. 3b) at the phase
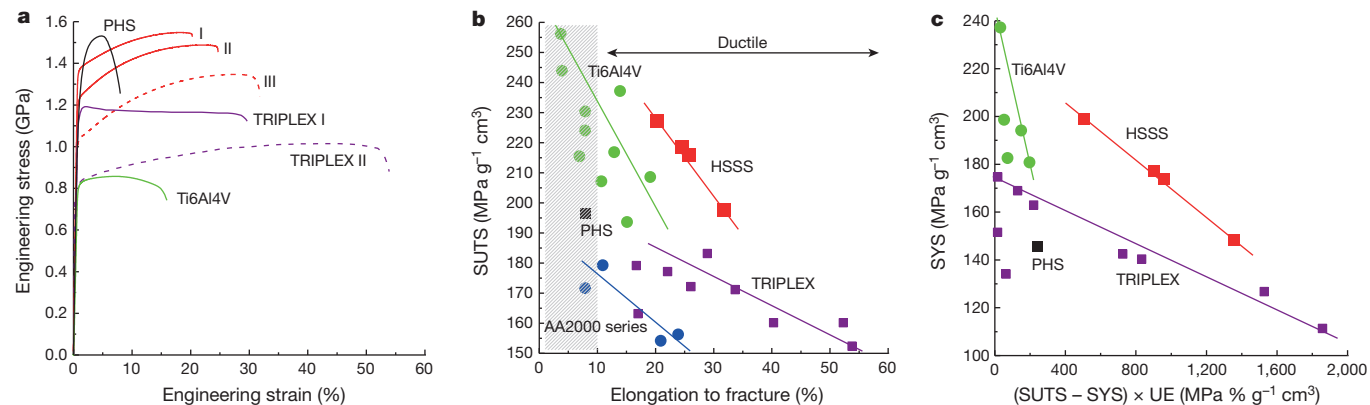
Figure 2 | Room-temperature tensile properties of HSSS compared with selected metallic alloys of high specific strength[6,7,22–27]. See Extended Data Tables 1 and 2. **a**, Engineering stress–strain curves of the annealed HSSS (I–III) (see Extended Data Table 2 for details) compared with κ-carbide-strengthened TRIPLEX steels[6,7], a commercial titanium alloy and a fully martensitic press hardening steel (PHS). **b**, Specific ultimate tensile strength (SUTS) versus

elongation to fracture of HSSS compared with press hardening steel and the available literature data[6,7,22–27] (AA2000 is a commercial aluminium alloy). **c**, Specific yield strength (SYS) versus density-compensated tensile toughness increase during uniform elongation (UE) of HSSS compared with selected metallic alloys of high specific strength[6,7,23–25].
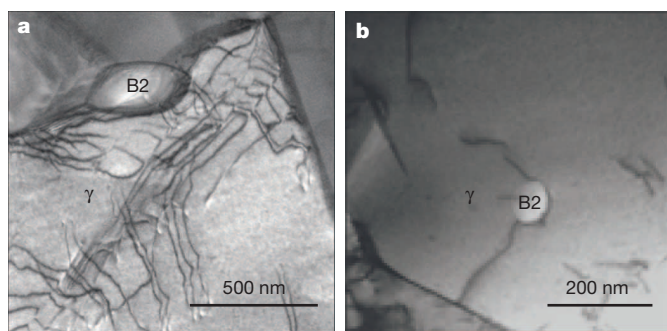
Figure 3 | Scanning TEM images of HSSS after tensile deformation (0.5% strained) showing interaction of dislocations and B2 particles. The non-shearable nature of B2 particles is depicted. **a**, Dislocations pile up at the interface of B2 particle and austenite (γ) matrix. **b**, A dislocation bows out at the B2/γ interface.

interface. It is clear that B2 particles are not sheared by gliding dislocations. The non-shearable nature of the B2 particle accounts for the high work hardening rate even at ultrahigh yield strength levels of over 1 GPa, leading to ductile ultrahigh-specific-strength steels.

These findings provide a new alloy-design route to lightweight steels, demonstrating that the combination of specific strength and ductility accessible to steels is greater than previously thought, and increasing the density-compensated tensile damage tolerance of structural metal for terrestrial applications. Furthermore, the attractive combination of physical and mechanical properties in the low-density steel described here is obtainable by simple thermal treatments which are compatible with existing commercial processes of the steel industry (Extended Data Table 2). The tuning of the distribution and morphology of brittle intermetallic compounds in steel matrix may be useful in many other steel applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Davis, S., Diegel, S. & Boundy, R. *Transportation Energy Data Book* 32nd edn, 4–17 (Oak Ridge National Laboratory, 2013).
2. Militzer, M. Materials science: a synchrotron look at steel. *Science* **298,** 975–976 (2002).
3. Kim, H., Suh, D. W. & Kim, N. J. Fe-Al-Mn-C lightweight structural alloys: a review on the microstructures and mechanical properties. *Sci. Technol. Adv. Mater.* **14,** 014205 (2013).
4. Suh, D. W. & Kim, N. J. Low-density steels. *Scr. Mater.* **68,** 337–338 (2013).
5. Zargaran, A., Kim, H. S., Kwak, J. H. & Kim, N. J. Effects of Nb and C additions on the microstructure and tensile properties of lightweight ferritic Fe–8Al–5Mn alloy. *Scr. Mater.* **89,** 37–40 (2014).
6. Sutou, Y., Kamiya, N., Umino, R., Ohnuma, I. & Ishida, K. High-strength Fe-20Mn-Al-C-based alloys with low density. *ISIJ Int.* **50,** 893–899 (2010).
7. Frommeyer, G. & Brüx, U. Microstructures and mechanical properties of high-strength Fe-Mn-Al-C light-weight TRIPLEX steels. *Steel Res. Int.* **77,** 627–633 (2006).
8. Gutierrez-Urrutia, I. & Raabe, D. Influence of Al content and precipitation state on the mechanical behavior of austenitic high-Mn low-density steels. *Scr. Mater.* **68,** 343–347 (2013).
9. Gschneidner, K. Jr *et al.* A family of ductile intermetallic compounds. *Nature Mater.* **2,** 587–591 (2003).
10. Kuc, D., Niewielski, G. & Bednarczyk, I. Structure and plasticity in hot deformed FeAl intermetallic phase base alloy. *Mater. Charact.* **60,** 1185–1189 (2009).
11. Stoloff, N. S., Liu, C. T. & Deevi, S. C. Emerging applications of intermetallics. *Intermetallics* **8,** 1313–1320 (2000).
12. Liu, C. T., Stringer, J., Mundy, J. N., Horton, L. L. & Angelini, P. Ordered intermetallic alloys: an assessment. *Intermetallics* **5,** 579–596 (1997).
13. Cao, G. H. *et al.* Determination of slip systems and their relation to the high ductility and fracture toughness of the B2 DyCu intermetallic compound. *Acta Mater.* **55,** 3765–3770 (2007).
14. Davies, R. G. Influence of martensite composition and content on the properties of dual phase steels. *Metall. Trans. A* **9,** 671–679 (1978).
15. Thomser, C., Uthaisangsuk, V. & Bleck, W. Influence of martensite distribution on the mechanical properties of dual phase steels: experiments and simulation. *Steel Res. Int.* **80,** 582–587 (2009).
16. Choi, K. *et al.* Effect of aging on the microstructure and deformation behavior of austenite base lightweight Fe-28Mn-9Al-0.8C steel. *Scr. Mater.* **63,** 1028–1031 (2010).
17. Yoo, J. D. & Park, K. T. Microband-induced plasticity in a high Mn-Al-C light steel. *Mater. Sci. Eng. A* **496,** 417–424 (2008).
18. Breuer, J., Grün, A., Sommer, F. & Mittemeijer, E. J. Enthalpy of formation of B2-$Fe_{1-x}Al_x$ and B2-$(Ni,Fe)_{1-x}Al_x$. *Metall. Mater. Trans. B* **32,** 913–918 (2001).
19. Chumak, I., Richter, K. W. & Ipser, H. Isothermal sections in the (Fe, Ni)-rich part of the Fe-Ni-Al phase diagram. *J. Phase Equil. Diffus.* **29,** 300–304 (2008).
20. Leslie, W. C. *The Physical Metallurgy of Steels* 43–59 (TechBooks, 1981).
21. Richeton, T., Weiss, J. & Louchet, F. Breakdown of avalanche critical behaviour in polycrystalline plasticity. *Nature Mater.* **4,** 465–469 (2005).
22. Gil, F. J., Manero, J. M., Ginebra, M. P. & Planell, J. A. The effect of cooling rate on the cyclic deformation of β-annealed Ti-6Al-4V. *Mater. Sci. Eng. A* **349,** 150–155 (2003).
23. Tähtinen, S., Moilanen, P., Singh, B. N. & Edwards, D. J. Tensile and fracture toughness properties of unirradiated and neutron irradiated titanium alloys. *J. Nucl. Mater.* **307–311,** 416–420 (2002).
24. Marmy, P. & Leguey, T. Impact of irradiation on the tensile and fatigue properties of two titanium alloys. *J. Nucl. Mater.* **296,** 155–164 (2001).
25. Bowen, A. W. The influence of crystallographic orientation on tensile behaviour in strongly textured Ti6Al4V. *Mater. Sci. Eng.* **40,** 31–47 (1979).
26. Dursun, T. & Soutis, C. Recent developments in advanced aircraft aluminium alloys. *Mater. Des.* **56,** 862–871 (2014).
27. Merati, A. in *Corrosion Fatigue and Environmentally Assisted Cracking in Aging Military Vehicles (RTO-AG-AVT-140)* Ch. 24 (Research and Technology Organization, NATO, 2011).
28. Fan, D. W., Kim, H. S. & De Cooman, B. C. A review of the physical metallurgy related to the hot press forming of advanced high strength steel. *Steel Res. Int.* **80,** 241–248 (2009).

Author Contributions N.J.K. and H.K. designed the study; S.-H.K. and H.K. performed the research; S.-H.K., N.J.K. and H.K. analysed the data; and N.J.K and H.K. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.K. (hansoo-kim@postech.ac.kr).
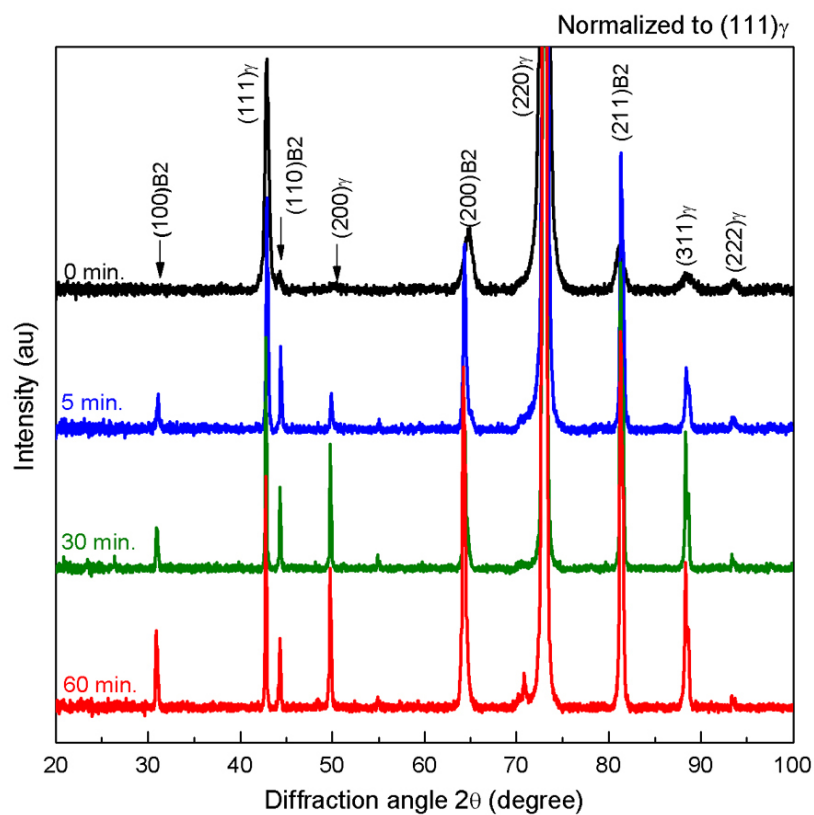
Extended Data Figure 1 | Change of X-ray diffraction pattern during annealing at 900 °C. With the increase in annealing time, the (100) peak of B2 becomes more pronounced owing to its precipitation (samples were water-quenched after annealing). au, arbitrary units.

Extended Data Table 1 | Composition of the present HSSS and reference materials

Material	Chemical composition, weight%											Comment
	Fe	C	Si	Mn	Al	Ti	Nb	V	Cr	Ni	B	
HSSS	Bal.	0.86	0.02	16.1	9.6	0.042	0.004	-	-	4.9	-	-
PHS	Bal.	0.22	0.24	1.2	-	0.04	-	-	0.2	-	0.0022	-
Ti6Al4V	0.12	0.01	-	-	6.1	Bal.	-	3.9	-	-	-	-
TRIPLEX I	Bal.	1.8	-	20	11	-	-	-	5	-	-	Ref. 6
TRIPLEX II	Bal.	1.0	-	28	12	-	-	-	-	-	-	Ref. 7

HSSS, high-specific-strength-steel; PHS, press hardening steel; Bal., balance; '-', not applicable.

Extended Data Table 2 | Thermal treatment condition and properties of the present HSSS and reference materials

Materials	Thermal treatment	Density compensated mechanical properties					Density	Comment
		SYS	SUTS	UE	TE	(SUTS-SYS) ×UE		
	Temp.×time; cooling	MPa/gcm⁻³		%		J/g	gcm⁻³	
HSSS	900 °C×2 m; 30 °C/s	199	227	18.0	20.3	508	6.82	I (Fig. 2a)
	900 °C×8 m; WQ	177	218	22.0	24.7	904		II (Fig. 2a)
	900 °C×15 m; WQ	148	198	27.6	31.8	1361		III (Fig. 2a)
	870 °C×15 m; WQ	174	216	22.7	25.6	960		-
PHS	900 °C×10 m; 35 °C/s	145	196	4.8	8.0	245	7.80	-
Ti6Al4V	950 °C×1 h; FC	182	194	6.8	15.2	75	4.43	Extended Data Table 1
	950 °C×1 h; WQ	237	256	1.8	3.8	34		
	700 °C×2 h; AC	223	237	-	14	-		Ref. 22
	1050 °C×1 h; 0.23 °C/s	185	215	-	7	-		
	1050 °C×1 h; 0.81 °C/s	194	224	-	8	-		
	1050 °C×1 h; 3.40 °C/s	201	230	-	8	-		
	1050 °C×1 h; 5.10 °C/s	202	244	-	4	-		
	730 °C×1.5 h; AC	194	217	6.7	13	153		Ref. 23
	730 °C×1.5 h; AC	181	208	7.2	19.2	200		Ref. 24
	-	198	207	6.6	10.8	57		Ref. 25
Al alloy 2000 series	-	117	154	-	21	-	2.77–2.78	Ref. 26
	-	132	179	-	11	-		
	-	134	172	-	8	-		
	-	110	156	-	24	-		
TRIPLEX	1100 °C×15 m; WQ	134	163	2.3	17.1	67	6.72	Ref. 6
	1100 °C×15 m; AC	151	160	2.2	40.4	19	6.64	
	1100 °C×15 m; WQ	175	183	2.3	29.0	20	6.51	
	1100 °C×15 m; WQ	142	177	21.0	22.2	729	6.43	
	1000 °C×15 m; WQ	127	160	45.7	52.4	1531	6.64	
	1000 °C×15 m; AC	163	171	26.9	33.8	223	6.64	
	900 °C×15 m; WQ	140	172	26.2	26.2	836	6.64	
	900 °C×15 m; AC	169	179	12.9	16.8	133	6.64	
	1050 °C×25 m; WQ	111	152	45.3	53.9	1862	6.57	Ref. 7

WQ, water quenching; AC, air cooling; FC, furnace cooling; '-', not available.

LETTER

Silylation of C–H bonds in aromatic heterocycles by an Earth–abundant metal catalyst

Anton A. Toutov[1]*, Wen-Bo Liu[1]*, Kerry N. Betz[1], Alexey Fedorov[1]†, Brian M. Stoltz[1] & Robert H. Grubbs[1]

Heteroaromatic compounds containing carbon–silicon (C–Si) bonds are of great interest in the fields of organic electronics and photonics[1], drug discovery[2], nuclear medicine[3] and complex molecule synthesis[4–6], because these compounds have very useful physicochemical properties. Many of the methods now used to construct heteroaromatic C–Si bonds involve stoichiometric reactions between heteroaryl organometallic species and silicon electrophiles[6,7] or direct, transition-metal-catalysed intermolecular carbon–hydrogen (C–H) silylation using rhodium or iridium complexes in the presence of excess hydrogen acceptors[8,9]. Both approaches are useful, but their limitations include functional group incompatibility, narrow scope of application, high cost and low availability of the catalysts, and unproven scalability. For this reason, a new and general catalytic approach to heteroaromatic C–Si bond construction that avoids such limitations is highly desirable. Here we report an example of cross-dehydrogenative heteroaromatic C–H functionalization catalysed by an Earth-abundant alkali metal species. We found that readily available and inexpensive potassium tert-butoxide catalyses the direct silylation of aromatic heterocycles with hydrosilanes, furnishing heteroarylsilanes in a single step. The silylation proceeds under mild conditions, in the absence of hydrogen acceptors, ligands or additives, and is scalable to greater than 100 grams under optionally solvent-free conditions. Substrate classes that are difficult to activate with precious metal catalysts are silylated in good yield and with excellent regioselectivity. The derived heteroarylsilane products readily engage in versatile transformations enabling new synthetic strategies for heteroaromatic elaboration, and are useful in their own right in pharmaceutical and materials science applications.

Heteroarylsilanes are important motifs in medicinal chemistry and drug discovery[2,3], advanced materials and polymer synthesis[1,10], and various biomedical applications[3,11]. In addition, they are emerging as one of the most versatile heteroaryl metal species for complex molecule synthesis owing to the high natural abundance and low toxicity of silicon[4–6]. At present, the most common approach to heteroaromatic C–Si bond construction involves the interception of heteroaryl lithium or magnesium reagents with silicon electrophiles (Fig. 1a, route A). However, this method is often limited in scope and requires prefunctionalization of heteroarenes by using pyrophoric organometallic species in stoichiometric quantities[7]. Powerful heteroaromatic functionalization strategies, such as Minisci-type radical substitutions[12] and Friedel–Crafts reactions[13,14], have been of limited use for C–Si bond construction owing to the difficulty of generating the corresponding silyl radicals and silylium ions. An efficient and regioselective sila-Friedel–Crafts reaction to access C3-silylated indoles was recently described[15], although the catalysis required a precious metal ruthenium (Ru) species for activation of the hydrosilane.

Thus far, only complexes based on precious metal elements, namely rhodium (Rh) and iridium (Ir), have been demonstrated to catalyse the intermolecular C–H silylation of heteroarenes with hydrosilanes[16] (Fig. 1a, route B). Two examples have been reported: an [Ir(OMe)(cod)]₂

precatalyst with a 4,4′-dtby ligand, used at 80 °C (ref. 9), and a [Rh(coe)₂(OH)]₂ precatalyst in combination with a MeO-BIPHEP ligand[8] (Me, methyl; cod, 1,5-cyclooctadiene; 4,4′-dtby, 4,4′-di-tert-butyl-2,2′-dipyridyl; coe, cis-cyclooctene.) Excess quantities of a sacrificial hydrogen acceptor were necessary for catalyst turnover in these systems. Although these are both important silylation methods, they rely on catalysts derived from rare and expensive precious metals, which can be a significant limitation, particularly for large-scale syntheses. Moreover, substrates containing Lewis-basic nitrogen functionalities are notably absent in both reports, limiting the use of these methods in pharmaceutical science and other biomedical applications. Thus, the development of a general catalytic method for heteroaromatic C–Si bond formation remains a considerable challenge in the broader field of C–H functionalization. Here we report that inexpensive and commercially available

a Route A: stoichiometric organometallic reactions

X = H or Hal
1) R–Li(MgX′)
2) [Si]–LG

X = H
Ir or Rh catalyst
[Si]–H
Excess H₂ acceptor
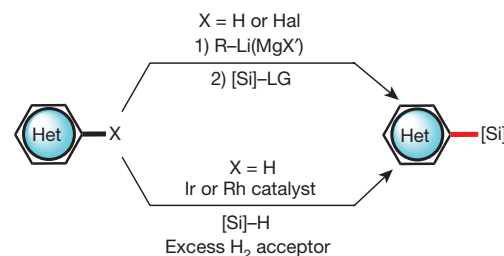
Route B: precious-metal-catalysed C–H silylation

b KOt-Bu-catalysed C–H silylation

1–20 mol% KOt-Bu
[Si]–H
–H₂ ↑
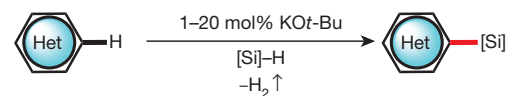
(Aza)indoles, furans, thiophenes, pyrroles, pyrazoles, etc.

- Earth-abundant metal (K) catalyst
- No H₂ acceptors or additives
- Mild reaction conditions
- Chemo- and regioselective
- TON up to 92
- >100 g scale

Figure 1 | Approaches to the silylation of heteroarenes. a, Route A, classical synthesis of heteroaryl silanes by reaction of organometallic species with silicon electrophiles. The organometallic species is typically prepared by deprotonation of heteroarenes or by lithium–halogen exchange of heteroaryl halides. X′ = Cl or Br; Hal, halogen; LG, leaving group. Route B, recently emerging direct, transition-metal-catalysed C–H activation/silylation. Excess amounts of hydrogen acceptors are required. **b**, A departure from the transition metal catalysis paradigm: KOt-Bu-catalysed, acceptorless, cross-dehydrogenative heteroaromatic C–H silylation with hydrosilanes. TON, turnover number.

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. †Present address: Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir Prelog Weg 2, CH-8093 Zürich, Switzerland.
*These authors contributed equally to this work.

potassium tert-butoxide (KOt-Bu) catalyses the acceptorless, cross-dehydrogenative coupling of aromatic heterocycles with hydrosilanes to generate heteroarylsilanes under mild conditions (Fig. 1b). The alkali metal catalyst is compatible with a range of functional groups including pyridines, piperidines and amines, making this novel C–H silylation method immediately applicable to medicinal chemistry and alkaloid natural product synthesis.

In a recent report, we described the reductive cleavage of C–O bonds in aryl ethers using stoichiometric quantities of alkali metal alkoxides to activate hydrosilanes at elevated temperatures[17]. We were surprised to observe minor by-products derived from *ortho*-silylation with dibenzofuran as the substrate. Considering prior reports of Lewis-base activation of hydrosilanes[18], we questioned whether these unanticipated silylation by-products could be pointing to a more general reaction manifold. Thus, with 1-methylindole as a model substrate, an extensive optimization exercise was conducted (Supplementary Information). We observed that the combination of a bulky basic anion (that is, Ot-Bu, trimethylsilanolate or bis(trimethylsilyl)amide) and, importantly, a potassium countercation led to the desired C–Si bond formation. KOt-Bu proved to be the ideal catalyst, furnishing synthetically useful C2-silylated indole **2a** in good yield and with >20:1 regioselectivity under mild conditions. The reaction can be optionally performed under solvent-free

conditions, which in certain cases leads to improved selectivity. Experiments and analyses to rule out catalysis by adventitious transition metal residues were carefully conducted (Supplementary Information).

A variety of indoles with Me, ethyl (Et), benzyl (Bn), phenyl (Ph) and the readily cleavable methoxylmethyl and 2-[(trimethylsilyl)ethoxy]methyl groups on nitrogen all lead to regioselective C2 silylation in moderate to good yields (Fig. 2, **2a–2f**). We then explored the influence of substituents at various positions of the indole nucleus and found that Me, OMe, OBn, CH_2OMe and Ph are all compatible, giving the desired products **2g–2n** in 48%–83% yield. Several hydrosilanes were examined and the silylation products (**2o–2x**) were obtained in good yield. A diverse range of N-, O- and S-containing heteroaromatics[19] (Fig. 3), including pyridine-containing scaffolds (**4a–4g** and **4j–4l**), undergo the reaction with high regioselectivity. Reactions at decreased catalyst loadings (1–3.5 mol%; **4j**, **4m** and **4n**) and on a large scale (**4h** and **4n**) demonstrate the robustness and preparative scale utility of the process. The reaction scaled to greater than 100 g without loss of catalyst activity under procedurally convenient conditions[20] (Fig. 4a).

In general, the reaction proved to be selective for electron-neutral and electron-rich heterocycles; indoles possessing electron-withdrawing groups are unreactive. To further probe the functional group tolerance of the method, a comprehensive robustness evaluation was performed[21].
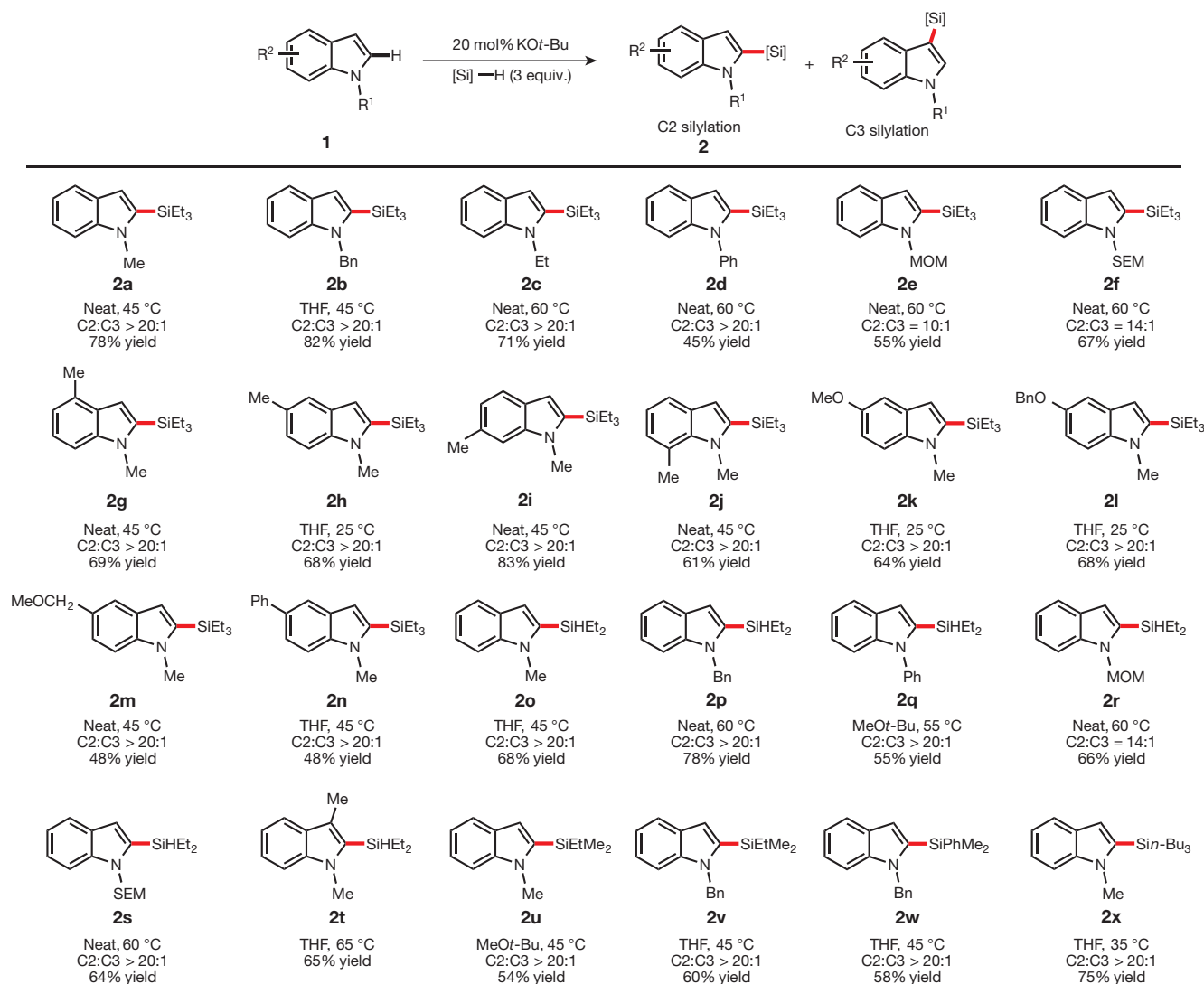
Figure 2 | Scope of the KOt-Bu-catalysed silylation of indoles. For the reactions of **2g** and **2i**, silylation on the benzylic methyl group was observed with tetrahydrofuran (THF) as solvent; solvent-free conditions often led to improved regioselectivity and yield. For the reaction of **2k**, silylation at C6 was observed as a by-product in THF. The reactions of 1,3-dimethyl indole with Et₃SiH and PhMe₂SiH were sluggish, probably owing to steric congestion at C2. For the reaction of **2o**, bisindolyldiethylsilane was isolated as a by-product. See Supplementary Information for details. [Si]–H = Et₃SiH, Et₂SiH₂, EtMe₂SiH, PhMe₂SiH or *n*-Bu₃SiH. MOM, methoxylmethyl; SEM, 2-[(trimethylsilyl)ethoxy]methyl.

20 mol% KOt-Bu

[Si]—H (3 equiv.)

Het—H **3** → Het—[Si] **4**

4a
THF, 45 °C
C2:C3 = 6:1
33% yield

4b
THF, 45 °C
C2:C3 > 20:1
31% yield

4c
THF, 45 °C
C2:C3 > 20:1
50% yield

4d
THF, 35 °C
C2:C3 > 20:1
71% yield

4e
THF, 45 °C
C2:C3 > 20:1
56% yield

4f
Neat, 60 °C
C2:C3 > 20:1
65% yield

4g
Neat, 60 °C
C2:C3 > 20:1
69% yield

4h
75 mmol scale
THF, 25 °C
C2:C3 > 20:1
93% yield
17.3 g

4i
THF, 25 °C
C2:C3 > 20:1
87% yield

4j
THF, 25 °C
α:β > 20:1
94% yield

4k
THF, 35 °C
α:β > 20:1
83% yield

4l
THF, 35 °C
α:β > 20:1
80% yield

4m
1 mol% KOt-Bu
5 mmol scale
THF, 25 °C
α:β > 20:1
92% yield, TON = 92

4n
73 mmol scale
THF, 25 °C
α:β > 20:1
95% yield
17.4 g

4o
THF, 25 °C
α:β > 20:1
78% yield

4p
THF, 25 °C
α:β > 20:1
88% yield

4q
THF, 25 °C
bis:mono > 20:1
82% yield

4r
THF, 25 °C
bis:mono > 20:1
73% yield

4s
THF, 25 °C
mono:bis = 10:1
80% yield

4t
THF, 25 °C
α:β > 20:1
71% yield

4u
Dioxane, 85 °C
o:p > 20:1
37% yield

4v
THF, 65 °C
o:p > 20:1
mono:bis = 15:1
71% yield

4w
THF, 65 °C
o:p > 20:1
64% yield

Figure 3 | KOt-Bu-catalysed silylation of N-, O- and S-containing heteroarenes. Multigram-scale syntheses were presented for **4h** and **4n**. Catalyst loadings can be reduced to 1 mol% with a TON of 92 (**4m**). For **4j**, with 3.5 mol% KOt-Bu, TON = 23 (82% yield); for **4n**, with 1.5 mol% KOt-Bu,

TON = 61 (91% yield). Bisfuranyldiethylsilane was isolated as a by-product in the reaction of **4o**. Unsubstituted thiophene and furan favoured 2,5-bis-silylation (**4q** and **4r**). See Supplementary Information for details. [Si]–H = Et₃SiH, Et₂SiH₂, EtMe₂SiH, PhMe₂SiH or n-Bu₃SiH.

The results showed that carbonyl groups in general are not tolerated, but are compatible if protected as the corresponding acetal. Ar–Br, Ar–I, Ar–CN, and Ar–NO₂ also shut down the reaction. However, Ar–F, Ar–Cl, Ar–CF₃, epoxide, N-alkyl aziridine, cis- and trans-olefins, acetylene, pyridine, and tertiary amine and phosphine moieties are all compatible with the silylation chemistry. Even free OH and NH groups are tolerated to some extent, apparently owing to a fortuitous silylative protection of the heteroatom in situ[18] (Supplementary Information).

Preliminary mechanistic investigations suggest the involvement of radical species. However, an elementary silyl radical generation-substitution mechanism seems to be unlikely owing to poor reactivity with electron-deficient heteroarenes[12,22]. Moreover, the rate of silylation is greater in sulphur-containing heteroarenes than in oxygen-containing heteroarenes, and is greater in oxygen-containing heteroarenes than in nitrogen-containing heteroarenes, as observed in an internal competition study, which provides complementary reactivity to electrophilic substitutions[14,15] and Minisci-type reactions[12,22]. These observations point to an underlying mechanism that is distinct from known heteroaromatic C–H functionalization reactions (Supplementary Information).

Heteroarylsilane derivatives are known to undergo a variety of powerful synthetic transformations; a number of representative examples are demonstrated here (Fig. 4b). For example, C2 Si-directed Suzuki–Miyaura cross-coupling by the method of Zhao and Snieckus[23], or Hiyama–Denmark cross-coupling[5] via heteroarylsilanol **6**[24], furnishes 2-arylated indole **5**. An unusual direct C7 functionalization of benzothiophene to give boronate esters **7** and **8** was achieved by using a blocking group strategy from silylated precursor **4h**[25].

Organosilicon has been extensively investigated in the development of advanced materials owing to silicon's unique physical and chemical properties[1,26]. To demonstrate the utility of our method in possible materials science applications, we prepared sila-heterocycle **9** in one step directly from the commercially available unfunctionalized heteroarene by an unprecedented double C–H functionalization involving intermolecular silylation followed by intramolecular silylation[10,16,27] (Fig. 4c). A high-yielding bis-silylation of thiophene oligomer **10** furnishes the starting material for an entirely transition-metal-free catalytic route to alternating copolymers[26]. Finally, the monoselective silylation of the 3,4-ethylenedioxythiophene monomer provides a potential strategy for the modification of polythiophene-derived materials (Fig. 4c, **11**).

Sila-drug analogues have garnered much attention from medicinal chemists because they can offer improved stability, solubility and pharmacokinetic properties compared with the parent all-carbon compounds[3]. Moreover, the installed organosilicon functionality can serve as a synthetic handle for subsequent elaboration, facilitating library synthesis and enabling structure–activity relationship studies. As a result, organosilicon-containing small molecules are of growing interest in pharmaceutical science, and the direct silylation of lead compounds would thus represent a new and potentially powerful tool in drug discovery[2]. To evaluate our method for such late-stage C–H functionalization applications, we subjected the antihistamine thenalidine and the antiplatelet drug ticlopidine to our catalytic silylation conditions. The reactions proceeded smoothly in the case of both active pharmaceutical ingredients, yielding the Si-containing target compounds **12** and **13a–c** in 56%–68% yield with excellent chemo- and regioselectivity (Fig. 4d). The

a Practical large-scale preparation of heteroarylsilane building blocks

b Si manipulation and blocking group strategies

Si-directed *ipso*-substitution/arylation

Hiyama–Denmark cross-coupling

Aerobic oxidation

Blocking group strategies

c Applications to functional materials

Inter/intra-molecular double C–H silylation

Mono C–H silylation

Multiple C–H silylations

d Late-stage, regioselective C–H silylation of pharmaceuticals

13a [Si] = SiEt₃, 57% yield
13b [Si] = SiHEt₂, 56% yield
13c [Si] = SiPhMe₂, 68% yield

e C(sp²)–H and C(sp³)–H silylation of arenes

Oxygen-directed C(sp²)–H silylation

Innate C(sp³)–H silylation

Figure 4 | Synthetic applications of the KOt-Bu-catalysed C–H silylation. a, Preparation of 142 g of C2-silylated indole building block **2a. b,** Application of heteroarylsilanes in cross-coupling and a formal C–H borylation at C7 of benzothiophene. **c,** Synthesis of precursors to advanced materials and polymers. **d,** Late-stage chemo- and regioselective modification of active pharmaceutical ingredients. **e,** KOt-Bu-catalysed functionalization of arenes by oxygen-directed sp², and innate benzylic sp³ C–H silylation. See Supplementary Information for details. [Si] = Et₃Si; *i*-Pr, isopropyl; dba, dibenzylideneacetone; Bpin, 4,4,5,5-tetramethyl-1,3,2-dioxaborolane; TMEDA, tetramethylethylenediamine; EDOT, 3,4-ethylenedioxythiophene.

piperidines, aniline, benzylic C–H bonds and aryl chloride moieties were all tolerated without any observed side reactions. Silylation of aza-analogue **14** also proceeded well, demonstrating the compatibility of our method with pyridine-containing complex molecules of potential pharmaceutical importance.

Finally, during our investigations we had observed minor amounts of sp² and sp³ C–H silylation by-products at ambient temperature in the cases of methoxy- and methyl-substituted indoles, respectively (that is, **15** and **16**; Fig. 4e). This led us to consider whether simple arenes would react analogously. The *ortho*-silylation of anisole[28] and the directing-group-free C(sp³)–H silylation of toluene[29,30] were discovered, furnishing silylated derivatives **17a** and **18a**, respectively. Four additional examples were demonstrated, providing silylarenes (**17b** and **17c**) and benzylsilanes (**18b** and **18c**) with excellent selectivity. Of particular note is the C(sp³)–H silylation of 2,6-lutidine, providing an example of C–H

silylation in an electron-deficient system. Interestingly, methoxy toluene **19** and benzyl ether **21**, both containing potentially reactive sp² and sp³ C–H bonds, are silylated with opposite selectivities to yield **20** and **22**. In the case of **22**, the reaction introduces a Si-substituted chiral centre. Optimization and further elaboration of these substrate classes is currently ongoing.

We have reported sp² and sp³ C–H silylation reactions catalysed by KOt-Bu, which is abundant, inexpensive, commercially available and bench stable. The transformation has been applied to an array of privileged heteroaromatic scaffolds and to a number of carbocyclic aromatic moieties. The potential for late-stage functionalization has been demonstrated by the direct silylation of active pharmaceutical ingredients. The extension of this work to non-aromatic systems is in progress, and detailed mechanistic investigations by experimental and computational methods are under way.

1. Zhang, F., Wu, D., Xu, Y. & Feng, X. Thiophene-based conjugated oligomers for organic solar cells. *J. Mater. Chem.* **21,** 17590–17600 (2011).
2. Showell, G. A. & Mills, J. S. Chemistry challenges in lead optimization: silicon isosteres in drug discovery. *Drug Discov. Today* **8,** 551–556 (2003).
3. Franz, A. K. & Wilson, S. O. Organosilicon molecules with medicinal applications. *J. Med. Chem.* **56,** 388–405 (2013).
4. Ball, L. T., Lloyd-Jones, G. C. & Russell, C. A. Gold-catalyzed direct arylation. *Science* **337,** 1644–1648 (2012).
5. Denmark, S. E. & Baird, J. D. Palladium-catalyzed cross-coupling reactions of silanolates: a paradigm shift in silicon-based cross-coupling reactions. *Chem. Eur. J.* **12,** 4954–4963 (2006).
6. Langkopf, E. & Schinzer, D. Uses of silicon-containing compounds in the synthesis of natural products. *Chem. Rev.* **95,** 1375–1408 (1995).
7. Whisler, M. C., MacNeil, S., Snieckus, V. & Beak, P. Beyond thermodynamic acidity: A perspective on the complex-induced proximity effect (CIPE) in deprotonation reactions. *Angew. Chem. Int. Ed.* **43,** 2206–2225 (2004).
8. Cheng, C. & Hartwig, J. F. Rhodium-catalyzed intermolecular C–H silylation of arenes with high steric regiocontrol. *Science* **343,** 853–857 (2014).
9. Lu, B. & Falck, J. R. Efficient iridium-catalyzed C–H functionalization/silylation of heteroarenes. *Angew. Chem. Int. Ed.* **47,** 7508–7510 (2008).
10. Tamao, K., Uchida, M., Izumizawa, T., Furukawa, K. & Yamaguchi, S. Silole derivatives as efficient electron transporting materials. *J. Am. Chem. Soc.* **118,** 11974–11975 (1996).
11. Ting, R., Adam, M. J., Ruth, T. J. & Perrin, D. M. Arylfluoroborates and alkylfluorosilicates as potential PET imaging agents: high-yielding aqueous biomolecular ^{18}F-labeling. *J. Am. Chem. Soc.* **127,** 13094–13095 (2005).
12. Du, W., Kaskar, B., Blumbergs, P., Subramanian, P.-K. & Curran, D. P. Semisynthesis of DB-67 and other silatecans from camptothecin by thiol-promoted addition of silyl radicals. *Bioorg. Med. Chem.* **11,** 451–458 (2003).
13. Furukawa, S., Kobayashi, J. & Kawashima, T. Development of a sila-Friedel–Crafts reaction and its application to the synthesis of dibenzosilole derivatives. *J. Am. Chem. Soc.* **131,** 14192–14193 (2009).
14. Curless, L. D., Clark, E. R., Dunsford, J. J. & Ingleson, M. J. E–H (E = R$_3$Si or H) bond activation by B(C$_6$F$_5$)$_3$ and heteroarenes; competitive dehydrosilylation, hydrosilylation and hydrogenation. *Chem. Commun.* **50,** 5270–5272 (2014).
15. Klare, H. F. T. *et al.* Cooperative catalytic activation of Si–H bonds by a polar Ru–S bond: regioselective low-temperature C–H silylation of indoles under neutral conditions by a Friedel-Crafts mechanism. *J. Am. Chem. Soc.* **133,** 3312–3315 (2011).
16. Seregin, I. V. & Gevorgyan, V. Direct transition metal-catalyzed functionalization of heteroaromatic compounds. *Chem. Soc. Rev.* **36,** 1173–1193 (2007).
17. Fedorov, A., Toutov, A. A., Swisher, N. A. & Grubbs, R. H. Lewis-base silane activation: from reductive cleavage of aryl ethers to selective *ortho*-silylation. *Chem. Sci.* **4,** 1640–1645 (2013).
18. Weickgenannt, A. & Oestreich, M. Potassium tert-butoxide-catalyzed dehydrogenative Si–O coupling: reactivity pattern and mechanism of an underappreciated alcohol protection. *Chem. Asian J.* **4,** 406–410 (2009).
19. Song, J. J. *et al.* Organometallic methods for the synthesis and functionalization of azaindoles. *Chem. Soc. Rev.* **36,** 1120–1132 (2007).
20. Li, C.-J. & Trost, B. M. Green chemistry for chemical synthesis. *Proc. Natl Acad. Sci. USA* **105,** 13197–13202 (2008).
21. Collins, K. D. & Glorius, F. A robustness screen for the rapid assessment of chemical reactions. *Nature Chem.* **5,** 597–601 (2013).
22. Seiple, I. B. *et al.* Direct C–H arylation of electron-deficient heterocycles with arylboronic acids. *J. Am. Chem. Soc.* **132,** 13194–13196 (2010).
23. Zhao, Z. & Snieckus, V. Directed *ortho* metalation-based methodology. Halo-, nitroso-, and boro-induced ipso-desilylation. Link to an in situ Suzuki reaction. *Org. Lett.* **7,** 2523–2526 (2005).
24. Lee, M., Ko, S. & Chang, S. Highly selective and practical hydrolytic oxidation of organosilanes to silanols catalyzed by a ruthenium complex. *J. Am. Chem. Soc.* **122,** 12011–12012 (2000).
25. Hansen, M. M. *et al.* Lithiated benzothiophenes and benzofurans require 2-silyl protection to avoid anion migration. *Synlett* **8,** 1351–1354 (2004).
26. Wang, Y. & Watson, M. D. Transition-metal-free synthesis of alternating thiophene-perfluoroarene copolymers. *J. Am. Chem. Soc.* **128,** 2536–2537 (2006).
27. Kuznetsov, A., Onishi, Y., Inamoto, Y. & Gevorgyan, Y. Fused heteroaromatic dihydrosiloles: synthesis and double-fold modification. *Org. Lett.* **15,** 2498–2501 (2013).
28. Oyamada, J., Nishiura, M. & Hou, Z. Scandium-catalyzed silylation of aromatic C–H bonds. *Angew. Chem. Int. Ed.* **50,** 10720–10723 (2011).
29. Kakiuchi, F., Tsuchiya, K., Matsumoto, M., Mizushima, E. & Chatani, N. Ru$_3$(CO)$_{12}$-catalyzed silylation of benzylic C–H bonds in arylpyridines and arylpyrazoles with hydrosilanes via C-H bond cleavage. *J. Am. Chem. Soc.* **126,** 12792–12793 (2004).
30. Sakakura, T., Tokunaga, Y., Sodeyama, T. & Tanaka, M. Catalytic C–H activation. Silylation of arenes with hydrosilane or disilane by RhCl(CO)(PMe$_3$)$_2$ under irradiation. *Chem. Lett.* **16,** 2375–2378 (1987).

Author Contributions A.A.T., W.-B.L and K.N.B. developed the reactions, performed the experiments and analysed data. A.F. analysed data. A.A.T and R.H.G. had the idea for and directed the investigations with W.-B.L and B.M.S. A.A.T. and W.-B.L. prepared the manuscript with contributions from all authors. All authors contributed to discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H.G. (rhg@caltech.edu) or B.M.S. (stoltz@caltech.edu).

LETTER

A seismic reflection image for the base of a tectonic plate

T. A. Stern[1], S. A. Henrys[2], D. Okaya[3], J. N. Louie[4], M. K. Savage[1], S. Lamb[1], H. Sato[5], R. Sutherland[1,2] & T. Iwasaki[5]

Plate tectonics successfully describes the surface of Earth as a mosaic of moving lithospheric plates. But it is not clear what happens at the base of the plates, the lithosphere–asthenosphere boundary (LAB). The LAB has been well imaged with converted teleseismic waves[1,2], whose 10–40-kilometre wavelength controls the structural resolution. Here we use explosion-generated seismic waves (of about 0.5-kilometre wavelength) to form a high-resolution image for the base of an oceanic plate that is subducting beneath North Island, New Zealand. Our 80-kilometre-wide image is based on P-wave reflections and shows an approximately 15° dipping, abrupt, seismic wave-speed transition (less than 1 kilometre thick) at a depth of about 100 kilometres. The boundary is parallel to the top of the plate and seismic attributes indicate a P-wave speed decrease of at least 8 ± 3 per cent across it. A parallel reflection event approximately 10 kilometres deeper shows that the decrease in P-wave speed is confined to a channel at the base of the plate, which we interpret as a sheared zone of ponded partial melts or volatiles. This is independent, high-resolution evidence for a low-viscosity channel at the LAB that decouples plates from mantle flow beneath, and allows plate tectonics to work.

The original concept of plate tectonics envisaged lithospheric plates that were mechanically and kinematically distinct from the underlying convecting mantle[3]. Subsequently, the base of the lithosphere has been defined as a critical isotherm[3], as a transition zone from conductive to convective heat flow, or as the base of an elastic layer capable of supporting flexural stresses on a geological timescale[4]. As these definitions are largely underpinned by thermal criteria, it is perhaps surprising that a marked change in geophysical properties at a depth of around 100 km is sufficiently consistent to define the LAB[5]. For example, a seafloor magnetotelluric (MT) study of young oceanic lithosphere identifies a thin (≤25 km thick) high-conductivity layer at the base of the lithosphere (at 45–70 km depth), which is interpreted as a melt-rich, low-viscosity channel[6]. Recent receiver-function studies have indicated that in places the LAB, at depths of 60–110 km, occurs where the S-wave speed (v_S) drops by ~8% over a depth range of less than 11 km (refs 1, 2). But the ability to resolve the sharpness of the LAB, which is critical for determining its nature[7], depends on the wavelength of the seismic waves used[1,7]. Active-source seismic methods allow a marked improvement in resolution, as large (>200 kg) dynamite shots can create reflected P waves with wavelengths of <1 km in the mantle[8,9], compared to wavelengths of 10–40 km for the teleseismic waves used to generate receiver functions[5]. Active-source and receiver-function methods do, however, provide a natural complement to each other because, whereas the former give vastly improved resolution, receiver functions provide crucial information on phase and therefore allow us to assess if there is a decrease or increase in seismic velocity across a converting interface.

Our data were acquired during the SAHKE (Seismic Array on the HiKurangi margin Experiment[10]) experiment (Fig. 1), which was located where the ~120-Myr-old[10] (120 Ma) Pacific plate and Hikurangi Plateau subduct westward beneath continental New Zealand (Fig. 1). We deployed 878 vertical and 300 three-component seismographs along a 85-km-long line[10], with a geophone spacing of 50–100 m. This design, together with 12 (500-kg) explosions in steel-cased boreholes, yielded high signal-to-noise ratios and deep P-wave reflections in a wide frequency band of 8–20 Hz (Fig. 2 and Extended Data Fig. 1).

Structure in the initial 12 s of two-way travel-time (TWTT) is seen on all shot records and shows the top of the subducted Pacific plate at depths of 15–30 km, which is concordant with Benioff Zone seismicity[10]
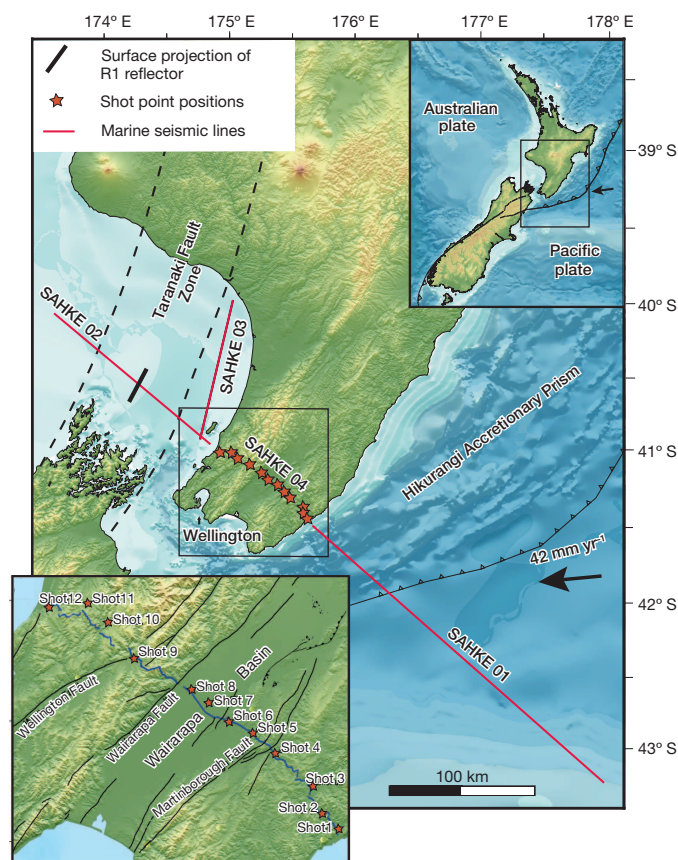
Figure 1 | Location map for SAHKE lines 01–04 and shot points for SAHKE04. Main panel, offshore lines SAHKE 01–03 are multichannel lines recorded both on land and offshore for analysis of crustal structure[10]. Red stars show position of 12 dynamite shots, of 500 kg each, on the 85-km-long onshore SAHKE04 line. Vertical component seismographs (887) were installed. The direction of oblique subduction of the Pacific plate beneath the Australian plate is shown by the black arrow, but the dip direction, or plunge, of the subducted plate is parallel to the SAHKE04 line. Top inset, plate tectonic setting of New Zealand. Bottom inset, enlarged section showing seismograph lines and shot positions of the SAHKE04 line. Mapped faults crossed by the seismic line are labelled.

[1]Institute of Geophysics, Victoria University, Salamanca Road, Wellington 6140, New Zealand. [2]Institute of Geological and Nuclear Sciences, 1 Fairway Drive, Lower Hutt 5010, New Zealand. [3]Department of Earth Sciences, University of Southern California, 3651 Trousdale Parkway, Los Angeles, California 90210, USA. [4]Seismological Observatory, University of Nevada, 1664 North Virginia Street, Reno, Nevada 90210, USA. [5]Earthquake Research Institute, Tokyo University, 1-1-1 Yoyoi, Tokyo 113-0032, Japan.
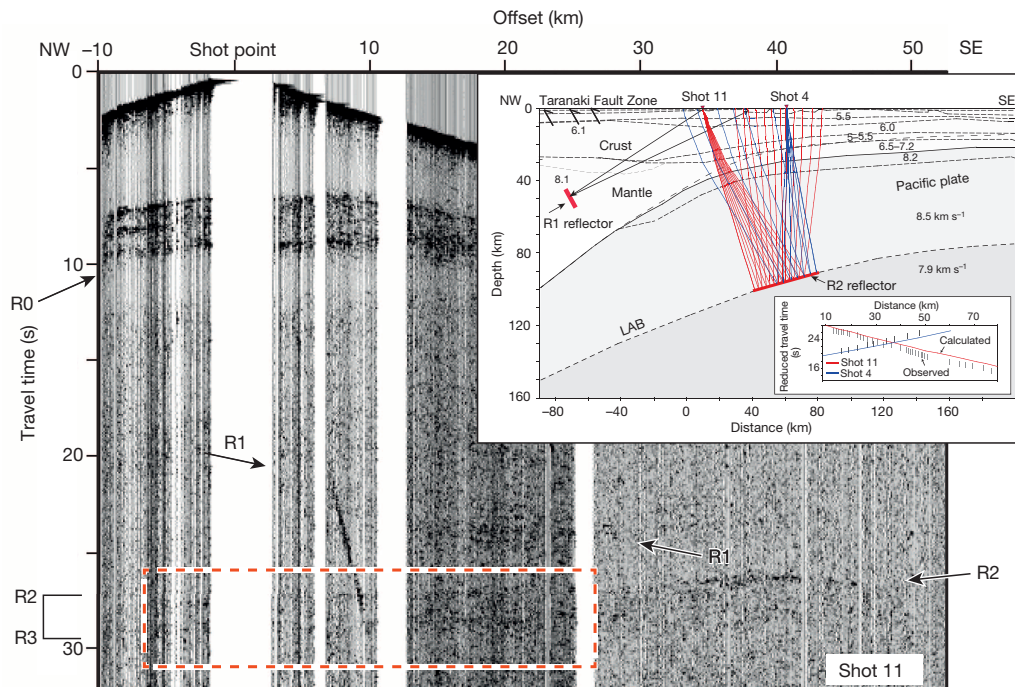
Figure 2 | Shot record for shot 11 and ray tracing model for arrivals from shots 11 and 4. Main panel, data shown have been band-pass filtered (12–20 Hz) and an attribute filter[16] applied. The left scale is in two way travel time (s). Between 7 and 10 s (~20–30 km) are three distinct reflections linked to under-plated sediment on top of the subducted plate[10]. R0 is an interpreted reflection from the top of the subducted plate, at the interface between under-plated sediment and oceanic crust[28]. Deeper events are the 20 s (at zero offset), southeast-dipping, R1 reflector and the 25–30 s doublet R2 and R3 reflectors that are interpreted as the LAB zone (dipping to northwest). The red dashed rectangular box outlines where a doublet (discussed in text) can be seen with a thickness of ~2 s two-way travel-time (TWTT). The steeply dipping event at ~8–10 km offset is the airwave. Inset: a ray-tracing solution

based on reduced (at 6 km s^{-1}) travel-time picks of the R2 reflector from shots 11 and 4 (secondary inset), and the calculated and observed travel time fit. The depth to the R2 reflector is 93 and 102 km beneath shots 4 and 11, respectively, based on a velocity model built from previous studies[10,12] (Extended Data Fig. 4b), and it dips 12°–15° to the northwest. The incidence angle range for the R2 reflections is ~0°–15°. The position of the southeast-dipping R1 reflector is schematically shown based on ray tracing and migration (Extended Data Fig. 2). Note that R1 is located along-strike but 50 km northwest off the end of the profile and beneath the Taranaki Fault Zone (Fig. 1). Numbers represent P-wave speeds in km s^{-1}. LAB is the interpreted lithosphere–asthenosphere boundary.

(Fig. 2). A strong reflector (R0 on Fig. 2) at about 9 s TWTT has been previously identified as the interface between a channel of low-v_P subducted sediments and the top of the Pacific plate. The Moho for the Pacific plate is seen at ~13 s TWTT, at zero offset, as a blurred zone of reflectivity that becomes prominent as a wide-angle reflection at an offset of ~55 km (Extended Data Figs 2 and 3).

Two deeper reflections observed between 20 and 30 s TWTT are seen on shot 11 (Fig. 2), and are referred to as R1 and R2. They dip steeply to the southeast, and gently to the northwest, respectively. Migration solutions (Fig. 2 inset, Extended Data Fig. 2b, c) indicate that the likely origin of the R1 reflection(s) is west of the SAHKE04 profile and within the mantle of the adjacent Australian plate. These R1 reflections are interpreted as originating from a deep thrust zone linked to a known period of Miocene tectonic shortening[11], and their geological interpretation is not discussed here. The key focus of this study is the R2 and associated R3 (2–3 s TWTT below R2) reflectors, which are at a depth of ~100 km (based on velocity models for the crust and mantle of southern North Island[10,12]). All 12 shots and 878 receivers are used to produce an image (Fig. 3) where up to 15 shot-to-receiver traces, which reflect from a common subsurface reflection point, are stacked together to form a single trace (Extended Data Figs 4 and 5).

The most plausible interpretation of the R2 and R3 reflectors is that they are sharp contrasts in physical properties associated with the LAB of the Pacific plate. This way, R2 and R3 parallel the observed dip on the top of the plate[11], and define a 73 ± 1 km thick lithosphere (Fig. 2, Extended Data Fig. 4c), consistent with a sub-solidus model for the oceanic LAB[7] and slightly thinner than ~100-Ma oceanic plates elsewhere in the western Pacific[2]. We eliminate the possibility that R2 and R3 are

artefacts related to earthquakes, because they are seen in energy returns from more than one shot. In addition, forward modelling shows that they cannot be explained as multiples of the shallower R0 and R1 crustal reflectors (Methods). Finally, we investigated the possibility that they are 'sideswipes' or reflections out of the plane of our transect (Methods); crucially the lack of coherent energy in an 'out-of-plane' shot gather, compared to the in-plane gather (Extended Data Figs 6–8), as well as the complete absence of evidence for features with the required dimensions and orientation in the crust or subducted slab either side of our transect, make it difficult to sustain a side-swipe interpretation. We note that between 35 and 40 s, there is seismic energy (labelled R4; see Fig. 3a) that shows some coherency and parallelism to the overlying R2 and R3 reflectors. Although the timing for this event is consistent with it being a S–P conversion from R2 (Extended Data Fig. 9), it is not well enough defined to make a confident identification.

The R2 reflections are in the frequency range of 8–20 Hz (Extended Data Fig. 1), consistent with frequencies of other reported deep mantle reflections from large dynamite shots[8,9]—such high frequencies can be explained by a travel path mainly in the high-Q mantle of the subducted Pacific plate[13]. For example, a spectral ratio analysis (Extended Data Fig. 1e) for the frequency spectra of the R0 and R2 reflections provides an estimate for Q_P of ~1,000 for the oceanic plate, similar to that found in other seismic studies[13]. Additional insight into the nature of the R2 reflector comes from its seismic attributes. As a rule of thumb, to observe a reflection, the normalized thickness of the acoustic-impedance transition zone, d/λ, will be less than 0.5, where d is the transition thickness and λ is the seismic wavelength[1,14,15] (Extended Data Fig. 7f). For R2, the reflected wave has a central frequency of ~14 Hz and a typical
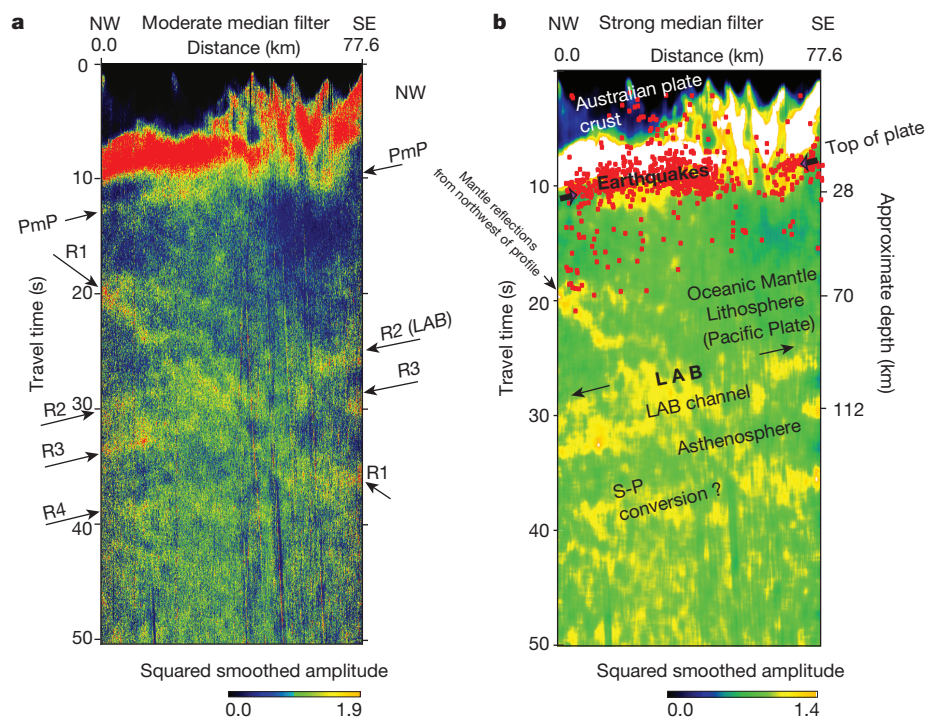
Figure 3 | A low-fold stack for the 12 shots recorded on 878 seismographs, with moderate and strong median filters. a, Low (1–15)-fold stack (fold = number of seismic traces that reflect from a common subsurface point that are stacked into a single trace) that has been processed with a median filter (Methods) that smooths traces in both space and time, and acts like a low-pass filter[29]. Plotted at approximate true scale with travel time on the left vertical axis, and approximate depth on the right vertical axis. The velocity model used and detailed fold geometry are given in Extended Data Fig. 4. The rectangular moving window smoothing kernel was over 21 traces (2.1 km) by 101 time samples (1.01 s). The R1 to R4 reflectors (see text) are highlighted. Colours are scaled to the amplitude of reflection. PmP is the interpreted position of the Moho for the Pacific plate. **b**, As for **a**, but with a stronger median filter: the rectangular moving window smoothing kernel was over 4.1 km laterally and by 2.01 s in travel time. Interpretations of the reflections labelled in **a** as R1 to R4 are shown as discussed in text. The southeast-dipping R1 mantle reflectors are interpreted to originate from a reflector off the northwest end of the SAHKE04 profile (see Extended Data Fig. 2). Earthquakes from the New Zealand GeoNet catalogue are plotted as red crosses, and show the top of the subducted Pacific plate between depths of ~15 km and 30 km along the SAHKE04 profile.

upper-mantle v_P of ~8 km s^{-1}, giving a thickness d of ~280 m. This implies that seismic velocity changes across R2 must occur over a thickness ≪1 km for a credible reflection to be observed.

Constraints on v_P below the R2 reflector are obtained by comparison of the instantaneous amplitude[16] of reflections with those for the shallower R0 reflector (Fig. 2), indicating a reflection coefficient[17] for

R2, C_r, of 0.065 ± 0.027, after corrections are made for intrinsic attenuation[16] within the mantle lithosphere and for geometrical spreading of the wave front (Methods). Deep-mantle reflections analysed in a similar manner from elsewhere in the world require reflection coefficients in the range 0.04 to 0.14 (refs 9, 18); our predicted value for C_r is at the lower end of this range. Using a value of 8.5 km s^{-1} for v_P at the base of
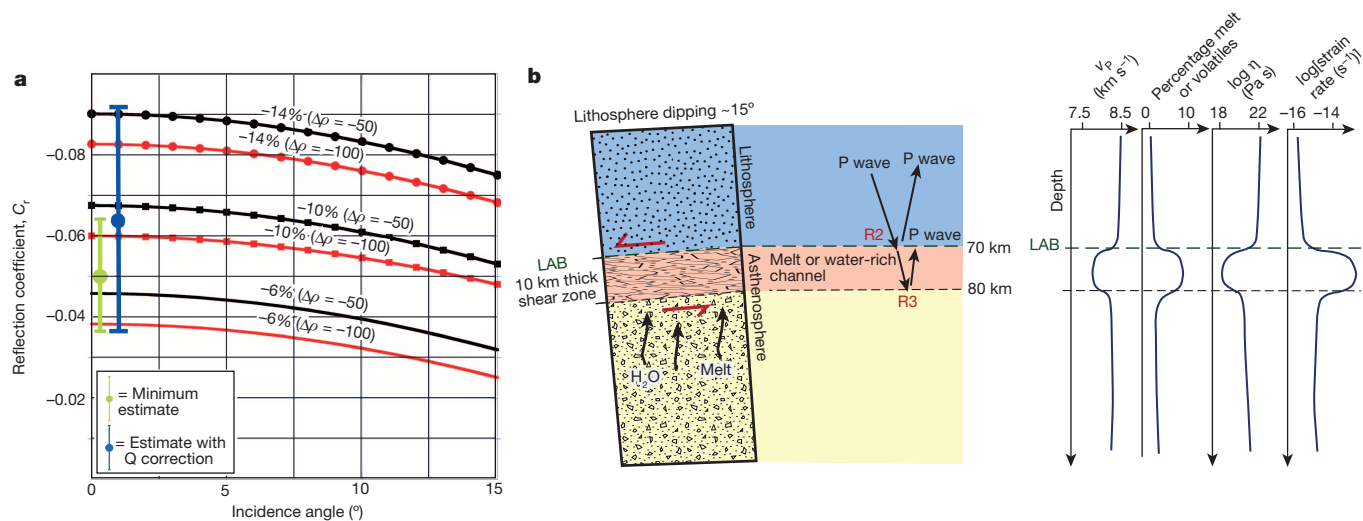
Figure 4 | Calculated reflection coefficient (C_r) for R2 for P–P reflections plotted against incidence angle, and interpretation carton of the LAB.
a, Plot of reflection coefficient versus incidence angle estimated for R2 and theoretical curves based on Zoeprittz equations[17]. v_P and density for the base of the lithosphere (R2) are fixed at 8.5 km s^{-1} and 3,400 kg m^{-3}, respectively, and v_P is allowed to vary below R2. The different curves show C_r values for labelled percentage drop in v_P and for two density contrasts of −50 and −100 kg m^{-3} across the LAB. We plot the magnitude of calculated reflection coefficient against a target region of $(C_r)_{R2} = 0.05 \pm 0.014$ for our estimate of the minimum value of C_r acquired from calibration with R0 (Methods). A velocity drop of $8 \pm 2\%$ is implied. If we make a correction for Q (Methods) for the different ray paths of R0 and R2, then $C_r = 0.065 \pm 0.027$ and the

estimate for a drop in v_P is $10 \pm 4\%$. The theoretical values of C_r are plotted against incidence angle out to 15° (with respect to a 15° dipping reflector), which is the maximum angle for stacked reflections based on our ray-tracing model. **b**, Summary schematic for our interpretation of physical properties at the base of the Pacific plate lithosphere. We interpret a channel about 10 km thick in which the v_P drops at least $8 \pm 2\%$ due to the pooling of melt[20,21] or water[7,22] at the LAB implying changes in: v_P from this study; percentage melt[20,21] as inferred from Δv_P, or percentage water content[7]; viscosity is schematic but based on our required melt level and model results[22]; and strain rate based on this study and assuming the background strain rate in the oceanic lithosphere is low.

the lithosphere[12], we find a v_P drop of $10 \pm 4\%$ below R2 is required to give the observed value of C_r (Fig. 4a), assuming an accompanying drop in density across R2 of between 50 and 100 kg m^{-3} (Methods). A positive impedance contrast here could also produce the R2 reflection, but would require $v_P > 9.2$ km s^{-1} below R2, which is not plausible at depths of 100 km (ref. 19). A drop in v_P of 8–10% also fits with receiver-function studies that show negative polarity and a similar percentage drop in v_S across the LAB[2].

A clear separation of ~2 s is observed between R2 and the underlying sub-parallel R3 reflectors for shot 11 (Fig. 2); on the stacked section (Fig. 3a) the separation between R2 and R3 is closer to 3 s due to smearing inherent in the stacking method (Methods). In addition, the amplitudes of R2 and R3 vary in a non-systematic way (Figs 2 and 3a), so that R3 cannot be a multiple of R2. We propose that R2 and R3 are the top and bottom of a low-seismic-wave-speed channel that is 8–12 km thick (that is, 2–3 s TWTT thick at a v_P of ~7.8 km s^{-1}).

Mechanisms for a marked drop in seismic wave speeds at the LAB include the onset of layered zones of partial melt[20,21], volatiles[7,22], and consequent anisotropy[2]. Partial melts of about 2% in a LAB channel are consistent with our estimated 8–10% drop in v_P (ref. 23), and is also in line with more general petrological arguments based on the study of xenoliths from oceanic volcanoes[20]. Whatever the mechanism, a special explanation is required to explain approximately equal, but opposite, acoustic impedance contrast at the boundaries of the channel. We propose that melt or volatiles, ponding at the base of the lithosphere, have created a low-viscosity[2,22] and high strain/strain-rate channel with distinct attributes (Fig. 4b), which is undergoing shear due to relative motion between the Pacific plate and underlying asthenospheric mantle (~9 cm yr^{-1}, in the hotspot reference frame[24]). If all this motion is accommodated in a zone ~10 km thick, it would produce strain rates of ~3×10^{-13} s^{-1}, and with finite shear strains of ~10 in ~1 Myr. We suggest that under these conditions there would be intense strain and melt localization[25], giving rise to the observed sharp acoustic impedance contrasts at the top and bottom of the layer.

A sheared, low-viscosity channel, up to 25 km thick, has been proposed at the LAB of a 20-Ma oceanic plate, seaward of the trench, based on a MT study[6]. Our study provides new and well-resolved evidence of a thinner channel extending down-dip of the trench beneath the 120-Ma subducted plate. These channels may therefore be a ubiquitous feature of oceanic plates, both inboard and outboard of the trench. An important question is then whether such channels also exist beneath continental lithosphere. In this regard, it may be significant that a double seismic reflection, similar to what we have observed, was found by a deep reflection survey at 22 s TWTT beneath the continental shelf of Norway, and has also been tentatively interpreted as the LAB[26]. This raises the intriguing possibility that thin channels of localized partial melt, or volatiles, are widespread features at the base of tectonic plates in general. If so, they would explain why tectonic plates apparently slide with minimal basal drag[27]. More high-resolution studies are required, however, and our experiment demonstrates it is possible to do this with active-source seismic methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Rychert, C. A., Rondenay, S. & Fischer, K. M. P-to-S and S-to-P imaging of a sharp lithosphere-asthenosphere boundary beneath eastern North America. *J. Geophys. Res.* **112**, B08314 (2007).
2. Kawakatsu, H. *et al.* Seismic evidence for sharp lithosphere-asthenosphere boundaries of oceanic plates. *Science* **324**, 499–502 (2009).
3. Parsons, B. & McKenzie, D. Mantle convection and the thermal structure of plates. *J. Geophys. Res.* **83**, 4485–4496 (1978).
4. Watts, A. B. *Isostasy and Flexure of the Lithosphere* (Cambridge Univ. Press, 2001).
5. Fischer, K. M., Ford, H. A., Abt, D. L. & Rychert, C. A. The lithosphere-asthenosphere boundary. *Annu. Rev. Earth Planet. Sci.* **38**, 551–575 (2010).
6. Naif, S., Key, K., Constable, S. & Evans, R. L. Melt-rich channel observed at the lithosphere–asthenosphere boundary. *Nature* **495**, 356–359 (2013).
7. Olugboji, T. M., Karato, S. & Park, J. Structures of the oceanic lithosphere-asthenosphere boundary: mineral-physics modeling and seismological signatures. *Geochem. Geophys. Geosyst.* **14**, 880–901 (2013).
8. Jarchow, C. M., Goodwin, E. B. & Catchings, R. D. Are large explosive sources applicable to resource exploration? *Leading Edge (Tulsa Okla.)* **9**, 12–17 (1990).
9. Steer, D. N., Knapp, J. H. & Brown, L. D. Super deep reflection profiling: exploring the continental mantle lid. *Tectonophysics* **286**, 111–121 (1998).
10. Henrys, S. A. *et al.* SAHKE geophysical transect reveals crustal and subduction zone structure at the southern Hikurangi margin, New Zealand. *Geochem. Geophys. Geosyst.* **14**, 2063–2083 (2013).
11. Lamb, S. H. Cenozoic tectonic evolution of the New Zealand plate-boundary zone: a paleomagnetic perspective. *Tectonophysics* **509**, 135–164 (2011).
12. Reyners, M., Eberhart-Phillips, D., Stuart, G. & Nishimura, Y. Imaging subduction from the trench to 300 km depth beneath the central North Island, New Zealand, with Vp and Vp/Vs. *Geophys. J. Int.* **165**, 565–583 (2006).
13. Eberhart-Phillips, D., Reyners, M., Chadwick, M. & Chiu, J.-M. Crustal heterogeneity and subduction processes: 3-D Vp, Vp/Vs and Q in the southern North Island, New Zealand. *Geophys. J. Int.* **162**, 270–288 (2005).
14. Warner, M. Free water and seismic reflectivity in the lower continental crust. *J. Geophys. Eng.* **1**, 88–101 (2004).
15. Bostock, M. G. Seismic waves converted from velocity gradient anomalies in the Earth's upper mantle. *Geophys. J. Int.* **138**, 747–756 (1999).
16. Zhou, H. *Practical Seismic Data Analysis* (Cambridge Univ. Press, 2014).
17. Sheriff, R. E. & Geldart, L. P. *Exploration Seismology* 2nd edn (Cambridge Univ. Press, 1995).
18. Warner, M. & McGeary, S. E. Seismic reflection coefficients from mantle fault zones. *Geophys. J. Int.* **89**, 223–230 (1987).
19. Hacker, B. R., Abers, G. A. & Peacock, S. M. Subduction factory 1. Theoretical mineralogy, densities, seismic wave speeds, and H$_2$O contents. *J. Geophys. Res.* **108** (B1), 2029 (2003).
20. Yamamoto, J., Korenaga, J., Hirano, N. & Kagi, H. Melt-rich lithosphere-asthenosphere boundary inferred from petit-spot volcanoes. *Geology* **42**, 967–970 (2014).
21. Sakamaki, T. *et al.* Ponded melt at the boundary between the lithosphere and asthenosphere. *Nature Geosci.* **6**, 1041–1044 (2013).
22. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
23. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity: effects of realistic partial melt geometries. *J. Geophys. Res.* **105**, 10975–10986 (2000).
24. Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150**, 321–361 (2002).
25. Hall, C. E. & Parmentier, E. M. Spontaneous melt localization in a deforming solid with viscosity variations due to water weakening. *Geophys. Res. Lett.* **27**, 9–12 (2000).
26. Lie, J., Pederson, T. & Husebye, E. S. Observations of seismic reflectors in the lower lithosphere beneath the Skagerrak. *Nature* **346**, 165–168 (1990).
27. Richardson, R. M. Ridge forces, absolute plate motions, and the intraplate stress field. *J. Geophys. Res.* **97**, 11739–11748 (1992).
28. Savage, M., Park, J. & Todd, H. Velocity and anisotropy structure of the Hikurangi subduction margin, New Zealand, from receiver functions. *Geophys. J. Int.* **168**, 1034–1050 (2007).
29. Duncan, G. & Beresford, G. Median filter behaviour with seismic data. *Geophys. Prospect.* **43**, 329–345 (1995).

Author Contributions All authors except J.N.L. and S.L. participated and led aspects of the data acquisition. J.N.L. developed and applied the median filter to produce Fig. 3a and Extended Data Fig. 5a, c. T.A.S. wrote the initial manuscript. S.A.H. organized the shots and permissions, and did processing for the pick migrations and the raw stack (Extended Data Figs 2 and 5b). D.O. and H.S. developed the initial shot gathers from the raw instrument data. T.A.S. and J.N.L. carried out numerical modelling (Fig.4b, Extended Data Figs 1d, 6 and 9). All authors discussed and commented on various versions of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.S. (tim.stern@vuw.ac.nz).

METHODS

Seismic processing. For the image in Fig. 3a we conducted a regular processing sequence with the following sequence using Global Claritas. (1) ADDGEOM: geometry and CDPs written to trace headers. (2) TREDIT: edit bad traces. (3) FDFILT: one filter per trace: bandpass (8-14-20-25 Hz). (4) AIRWAVE: air wave mute. (5) GENSORT: traces sorted into CDP bins with nominal spacing of 100 m. (6) NMO: normal moveout applied using input stacking velocity file (Extended Data Fig. 8). Stretch mute at 500% with taper of 1 s. (7) ATTRIBUTE: conversion to instantaneous amplitude. (8) STACK: stack with unity normalization; trace offset range = -50 to 50 km. (9) AGC: normal AGC. Window = 9,990 ms. (10) WRITESEGY: disk file written.

We applied to the final stack (Fig. 3a) an amplitude-enhancement smoothing method that is equivalent to a form of low-pass filter in both space and time. This is called a median filter[29], which is a nonlinear digital filtering technique used to remove noise. A median filter replaces each data point in a data set with the median of the data points in a specified region centred on that data point. The smoothing was applied to the stack after converting the seismogram traces, having both positive and negative amplitudes, to the all-positive instantaneous amplitude or envelope attribute[16]. The dimensions for the image in Fig. 3a of the rectangular moving window smoothing kernel were 21 traces (2.1 km) by 101 time samples (1.01 s). The smoothing centred the moving window over each sample in the stack, and placed at that location the square of the median of all the amplitude envelope samples within the window, equally weighted (a squared 'median filter' as used is in image processing as described at http://en.wikipedia.org/wiki/Median_filter). For the section in Extended Data Fig. 5c, the instantaneous amplitudes are multiplied at each sample by the squared median-filtered amplitudes of Fig. 3a. Thus, the stack is enhanced at each point by the squared median of the neighbouring amplitudes, making the distribution of amplitudes less Gaussian.

Calculating C_r for the R2 (LAB) reflector. If the absolute reflection coefficients (C_r) from a deep mantle reflection can be deduced from shot gathers, one can gain estimates for the change in seismic wave-speeds across the reflector.

The reflection coefficient at normal incidence is given by:

$$C_r = (Z_2 - Z_1)/(Z_1 + Z_2) \qquad (1)$$

where acoustic impedance $Z = \rho v_P$, v_P is P-wave speed, ρ is density and Z_1 and Z_2 are the associated values at top and bottom of a reflector. At non-zero incidence the S-wave velocities become important, and the Zoeprittz equations[17] give the general expression for C_r at any angle of incidence between $0°$ and $90°$. Density is also a parameter in equation (1), but as it usually only varies by only a few per cent its variation has a minor effect on C_r compared to the changes in seismic wave speeds. C_r can vary from -1 to 1. A negative value of C_r implies a phase change at the boundary. In our study the data are not clear enough to determine phase so we only consider the magnitude of C_r.

Calculating C_r of deep reflections often involves the calibration of the amplitude for the reflection of interest with a shallower one that is from an interface with a known acoustic impedance contrast[18]. For example, in marine experiments the amplitudes of deep mantle reflections recorded can be calibrated against the water bottom reflection[18]. As we are dealing with land-based data, we have to choose an upper crustal reflection as a calibration standard. In order to make this calibration between two reflections at different depths, two corrections must be made: for geometrical spreading of the wavefront, and for intrinsic attenuation between the two reflections (Q^{-1})[16,17]. These are discussed further below. Because of the uncertainties surrounding Q, and its nonlinear relationship to seismic amplitude, we can at least place a minimum bound on C_r if we ignore the effect of Q[18].

On shot 11 there is a sequence of three reflectors of varied quality between 7 and 10 s TWTT (Fig. 2). The third reflector (R0 on Fig. 2) is the strongest, and has been interpreted from the SAHKE active source experiment as the interface between underplated sediments and the underlying oceanic crust[10]. Receiver function studies[28] from the northwest end of the SAHKE line also clearly identify a low velocity layer on top of the subducted plate beneath southern North Island (Extended Data Fig. 9). Four studies identify v_P in this channel to be in the range 4.4–5.1 km s^{-1} with an average of 4.87 km s^{-1} and standard deviation of 0.32 km s^{-1} (refs 10, 28). We adopt $v_P = 4.87$ km s^{-1} and density = 2,500 kg m^{-3} for the low velocity channel, and $v_P = 6.5$ km s^{-1} (ref. 10) and density = 2,900 kg m^{-3} for the oceanic crust below R0, giving a reflection coefficient (based on equation (1)) of $(C_r)_{R0} = 0.22$. We assume the density uncertainty is small compared to that of the P-wave speeds. We allowed v_P of the oceanic crust below R0 to vary between 6.4 and 6.7 km s^{-1}, the v_P on top of R0 to vary between 4.4 and 5.1 km s^{-1}, and then ran a Monte Carlo simulation of 1,000 random samples to obtain a predicted value of the reflection coefficient of R0 ($\pm$2s.d.):

$$(C_r)_{R0} = 0.22 \pm 0.04$$

We solve for $(C_r)_{R2}$ where $(C_r)_{R0}/(C_r)_{R2} = (R0/R2)_{amp}$, and where $(R0/R2)_{amp}$ is the observed ratio for the R0 and R2 reflections.

To obtain an estimate of this ratio we compared and measured ratios of instantaneous amplitude, on common traces, for the near vertical incidence angle reflections R0 and R2, on shot 11. The only processing is a spherical divergence adjustment that compensates for the natural attenuation of signal due to geometrical spreading of the wavefront based on the adopted velocity model and travel time[17]. We made 13 independent measurements of $(R0/R2)_{amp}$ to give a mean value of 4.4 with standard deviation of 1.45. The standard error of the mean is therefore $1.45/(13)^{0.5} = 0.4$.

The predicted reflection coefficient of R2 is:

$$(C_r)_{R2} = (C_r)_{R0}/(R0/R2)_{amp} = 0.22/4.4 = 0.05 \qquad (2)$$

Uncertainties. We estimate the uncertainty from the propagation of errors for a quotient quantity[30]. That is, if $X = A/B$, the variance of X is given by

$$S_X^2 = (A^2 S_B^2 + B^2 S_A^2)/B^4$$

where S_B and S_A are standard deviations of the mean. So for $A = 0.22$ and $B = 4.4$, we find $S_A = 0.02$, $S_B = 0.4$, $A/B = 0.05$, and from the above, $S_X = 0.0064$. At the 95% confidence level (2σ): critical t (for 95% and $13 - 1 = 12$ degrees of freedom) is 2.18 (using Excel tinv(0.05,12) function). Therefore $(C_r)_{R2} = 0.05 \pm 2.18 \times 0.0064$ (95%) $= 0.05 \pm 0.014$.

We also applied a Monte Carlo method of error evaluation to estimate $(C_r)_{R2}$. We ran 1,000 random samples using the means and standard deviations of the two quantities that define $(C_r)_{R2}$ (that is, $(C_r)_{R0}$ and $(R0/R2)_{amp}$) and we get the following:

$$(C_r)_{R2} = 0.05 \pm 0.007 \text{ (1s.d.) or}$$

$$(C_r)_{R2} = 0.05 \pm 0.014 \text{ (2s.d.)} \qquad (3)$$

So identical results are obtained from both methods of error evaluation.

Finally, we consider a correction for intrinsic attenuation. The change in amplitude (ΔA) due to a seismic wave (frequency f) traversing a medium for a time t, with attenuation (Q^{-1}), is given by[16]:

$$\Delta A \approx \exp(-\pi f t/Q)$$

Now Q in the subducted plate beneath the North Island is high ($\sim$1,000), whereas Q in the crust above is much lower ($\sim$350) yet variable[13]. The central frequency of the R0 and R2 reflectors is about 19 and 14 Hz, respectively, and the two-way travel times are 9 and 27 s respectively.

The Q correction factor for R_{LAB} is then:

$$\Delta A \approx [\exp(-\pi f_{R0} t_{R0}/Q_{R0})/\exp(-\pi f_{R2} t_{R2}/Q_{R2\text{-avge}})] \qquad (4)$$

where $Q_{R2\text{-avge}}$ is the path averaged attenuation (over n layers) down to R2 and back given by[31]:

$$Q_{R2-\text{avge}} = \sum_{i=1}^{n} t_n / \sum (t_n/Q_n) \qquad (5)$$

where t_i and Q_i are respectively travel time and attenuation in the ith layer. We adopt a two layer model with t_1 and t_2 equal to 9 and 18 s, respectively, and with Q_1 and Q_2 equal to 350 and 1,000, respectively, and the path averaged Q to R2 is (from equation (5)) 617.

The correction factor ΔA (equation (4)) is sensitive to both adopted frequencies and Q values. For example, if we fix the central frequencies as above and adopt a range of $Q = 350 \pm 100$ for the crust based on published cross-sections[13], then the correction is 1.3 ± 0.2, which we use to scale the result of equation (3) to get:

$$(C_r)_{R2} = 0.065 \pm 0.027 \qquad (6)$$

Here we have calculated the uncertainty using the propagation of errors for a product[30]. This is the preferred value of $(C_r)_{R2}$ although the uncertainty has doubled from that in equation (3). Therefore, following ref. 18, an alternative, and more robust, conclusion is that if we ignore the correction for Q then the lower bound on $(C_r)_{R2}$ is $= 0.05 \pm 0.014$.

Tests for R2 being earthquakes, multiples or sideswipe. Earthquakes. R2 and R3 are seen on multiple shots (Extended Data Fig. 3) and form a coherent stacked section, even when the best shot (shot 11) is not included in the stack (Extended Data

Fig. 5a). We therefore reject the possibility that R2 and R3 are earthquakes that have fortuitously appeared across our records.

Multiples. Wave equation modelling (Extended Data Fig. 6) shows that multiples of reflections from the top of the plate reflection (R0 at 9 s) will not appear at 27 s with the correct moveout to replicate R2. The reason why such multiples, if they are present, will not be mistaken for the primary R2 and R3 reflections is that because the layers are dipping the first multiple will appear to be from a layer that is twice as deep as the reflector, and with a dip that is twice as large[32]. The next multiple after that will double in dip, and so on. For a dipping reflector the apex of the reflection hyperbola will move laterally by[17]:

$$X = 2d\sin\phi \qquad (7)$$

where ϕ and d are respectively the dip of, and perpendicular distance to, the reflector. The primary reflector for R* on Extended Data Fig. 6 has moved ~12 km updip from the shot point (that is, $d = 22$ km, $\phi = 15°$ in equation (7) above) and the multiple (M*1) has moved ~50 km updip ($d = 44$ km, $\phi = 30°$). So by the time the second multiple crosses the position of the R2 and R3 reflections, the multiples are dipping at a very high angle and could not be mistaken for R2 or R3 (Extended Data Fig. 6). Note that the synthetic seismic section generates much stronger multiples than are seen in reality due to effects of topography and near surface effects.

Sideswipe. Reflections from out of plane surfaces can appear on shot records looking like they are from reflectors directly below. We list five tests to check for sideswipe phenomena: (1) Coherency. We made shot gathers with a median filter for shot 11 that are both in line with the seismic line and at right angles to it (Extended Data Fig. 8). If R2 is sideswipe appearing with some coherence on the inline shot record, it should be even more coherent in the short cross line record. It is instead less coherent, suggesting R2 is more likely to be a reflection from a surface directly below the seismic line rather than one to a side. (2) Prestack depth migration. We performed pre-stack depth migration[16] both inline with the seismic profile and at right angles to it (cross line) (Extended Data Fig. 7). On the inline migration we see coherent energy captured in both the 30 and 90 km depth ranges, whereas there is no visible energy at all in the cross-line migration (Extended Data Fig. 7). This is consistent with the reflections coming from directly below the line, and not sideswipe from a structure to the side of the profile. (3) Frequency content of R1 and R2. We

showed from a spectral ratio analysis that the higher frequency of R2 (8–20 Hz) is consistent with a raypath from R0 to R2 through high (~1,000) Q upper mantle (Extended Data Fig. 1d). In contrast the R1 reflection, which has a similar travel time to R2, has a lower frequency content (6–10 Hz). When R1 is migrated (Extended Data Fig. 2) we can see its travel path is almost entirely through the crust, which in this area has much lower value of Q of about 300–400 (ref. 13). If R2 were indeed a sideswipe, its path would also be mainly through heavily fractured crust and we would then expect a lower frequency signal like R1. (4) Three-dimensionality in the survey geometry and effect on the stack. If we draw a straight line between shots 1 and 12 then there is deviation of the shot locations from that line of up to 8 km. These deviations would manifest themselves as large static corrections on a stack of sideswipe reflections, which would degrade the quality of the stack. (5) Strike of geology and structure. The dominant strike of the geological structure and strike slip faults in the vicinity of the SAHKE line is northeast–southwest, which is at right angles to the seismic line (Fig. 1). The strike-slip faults that have been active for the past 20 Myr are spaced at 10–20 km across the lower North Island; therefore, if there were structures in the lower crust, or upper mantle, at right angles to the faults they would have been differentially offset by motion on these faults and thus not likely to produce coherent returns over lateral distances of 10–20 km. Finally, no obvious offset or tear of the subducted plate is evident beneath the southern North Island[33].

30. Taylor, J. R. *An Introduction to Error Analysis* 2nd edn (University Science Books, 1997).
31. Menke, W., Levin, V. & Sethi, R. Seismic attenuation in the crust at the mid-Atlantic plate boundary in south-west Iceland. *Geophys. J. Int.* **122,** 175–182 (1995).
32. Badley, M. E. *Practical Seismic Interpretation* (International Human Resources Development Corporation, 1985).
33. Williams, C. A. *et al.* Revised interface geometry for the Hikurangi Subduction Zone, New Zealand. *Seismol. Res. Lett.* **84,** 1066–1073 (2013).
34. Seward, A. *et al. Seismic Array HiKurangi Experiment II (SAHKE II)* (Onshore Active Source Acquisition Report no. 2011/50, GNS Science, Lower Hutt, 2011).
35. Larsen, S. & Schultz, A. *ELAS3D:2D/3D Elastic Finite-difference Wave Propagation Code* (Lawrence Livermore National Laboratory Technical Report UCRL-MA-121792, 1995).

a/

Frequency spectrum

b/

Frequency spectrum

c/

Shot	Event	Frequency range (Hz)	Comment
11	R0	12-23	Best shot hole for generating high frequency energy. Drillers log: Blue greywacke – very firm from 27 m.
12	R0	7-18	Drill hole only reached basement in the last meter. Shot adjacent to strike slip fault.
11	R1	6-10	Shot 11 in very firm rock.
3	R2	4-15	Shot is in sediments.
11	R2	7-20	Shot 11 in very firm rock.
10	R2	8-14	Shot is in greywacke but near strike slip fault (Wellington Fault).
1	R4	3-8	Shot in sediments
12	R4	5-12	

d/

Spectral ratio

Extended Data Figure 1 | Frequency spectra and analysis. a, Frequency spectrum of the R2 reflection and the background noise. We only show frequencies above the 4.5-Hz cut-off frequency of the geophones used. **b**, Frequency spectrum of the R0 and R2 reflections. **c**, Table summarizing the frequency range for all events on the different shots; summaries of the geological conditions in each shot hole[34] are also given. **d**, A spectral ratio

analysis[16] between R2 and R0 reflections that yields a least squares linear fit to the data to give a gradient of -0.0627. Theoretically this gradient value can be equated to $\pi \Delta t / Q_P$ (ref. 16), where Δt is the TWTT between the R0 and R2 reflectors. For $\Delta t = 20$ s, we get an estimate of Q_P (inverse attenuation) of 1,002 with a standard error of ± 30 (based on least squares linear regression).
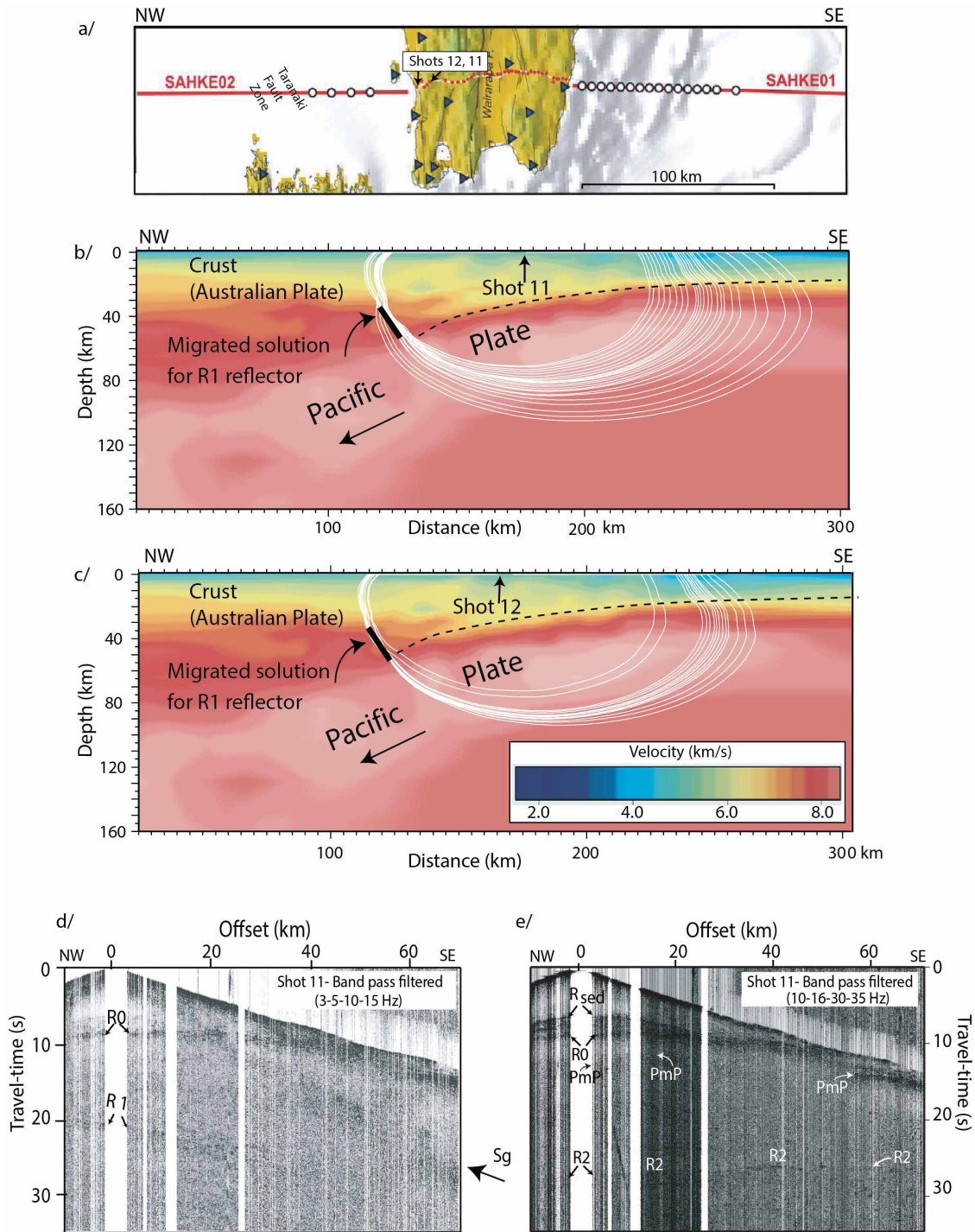
Extended Data Figure 2 | Pick migration for R1 reflector and shot gathers.
a, Plan view of onshore and offshore SAHKE lines with shots 11 and 12 labelled.
b, Pick migrations for R1 reflection for shot 11 based on a laterally varying velocity model, which is derived from earthquake and shot data[10]. We use the velocity model created by 3D tomography to carry out a migration of reflection picks[10]. Where the arcs converge to a constant solution gives the best structural interpretation of the reflector. We show the results for two shots (**b**, shot 11; **c**, shot 12) and the common solution. The solution (black bar) appears to be a reflector in the mantle of the Australian plate, which dips to the southeast and is located within the Taranaki Fault Zone. **d**, Shot gather for shot 11 with low band-pass filter of 5–10 Hz that brings out the low frequency, R1, southeast-dipping reflector (20 s depth at zero offset). It also shows the R0 reflector (~9 s at zero offset) for the top of the plate that dips to the northwest. Vertical axis is travel time (s) and horizontal axis is shot offset in km. S_g is the crustal refracted S wave. **e**, Shot gather for shot 11 with band pass filter 16–30 Hz that shows up the R2 reflector with its broad frequency content, and suppresses the low frequency R1 reflection. R_{sed} and P_mP represent reflections interpreted to come from the upper surface of the sediment channel on top of the oceanic crust, and from the Moho of the oceanic Pacific plate, respectively.
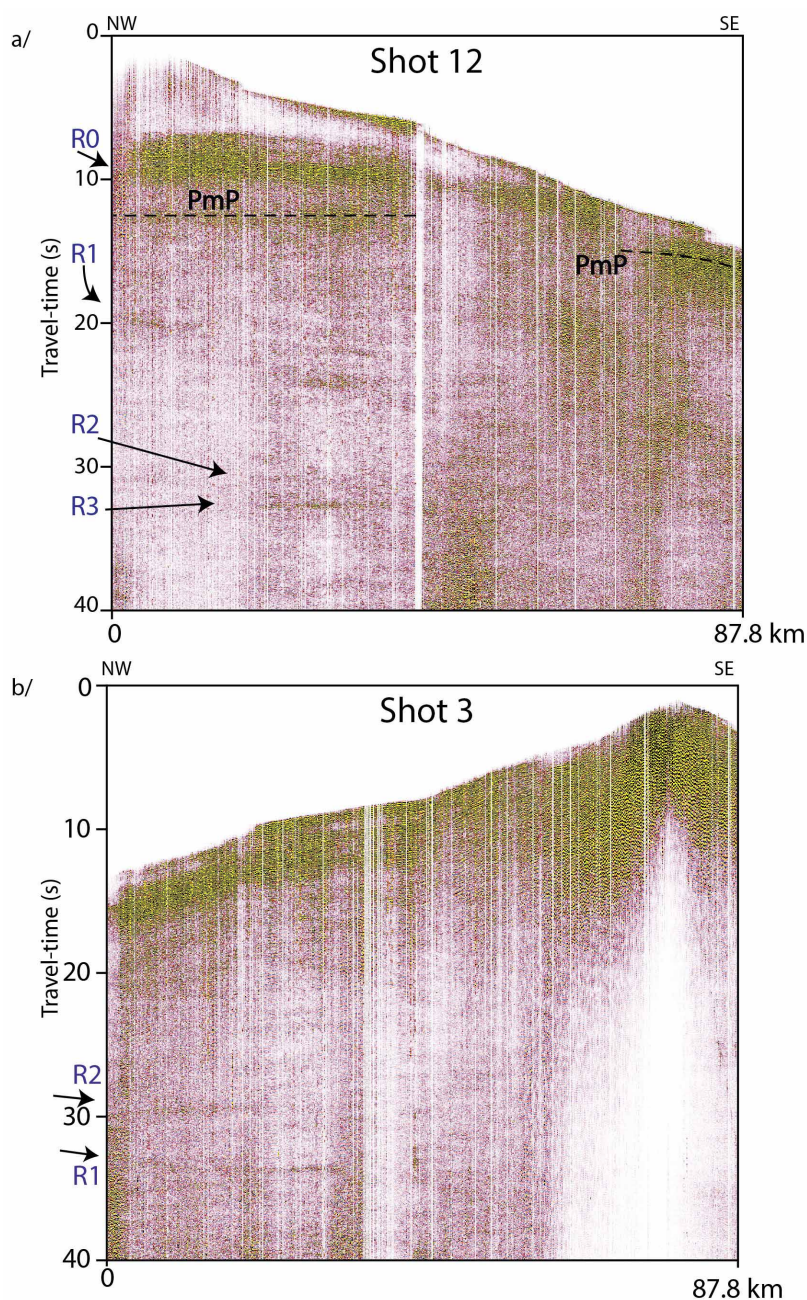
Extended Data Figure 3 | Two shot gathers from each line end showing reflections R0 to R3. a, Shot 12; **b**, shot 3. Both plots show data that have been band-passed between 8 and 25 Hz. The quality of the shots is highly dependent on the rock that the shot was placed in, with the drillers' logs[34] showing basement rock for shots 9, 10 and 11, and the very bottom of shot hole 12.

a/

b/

c/
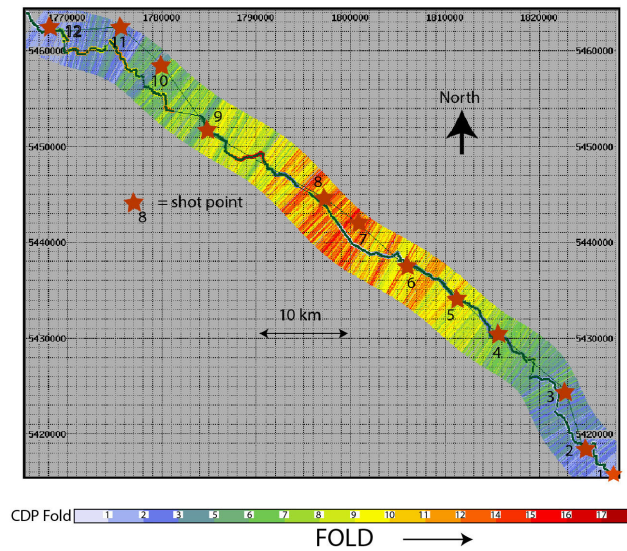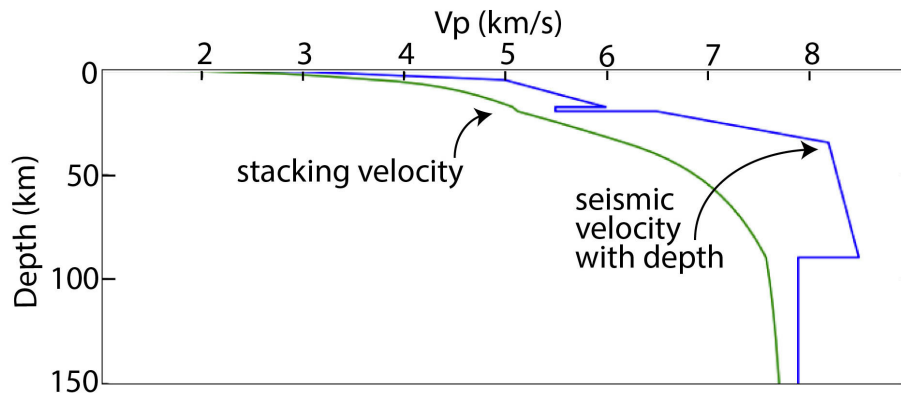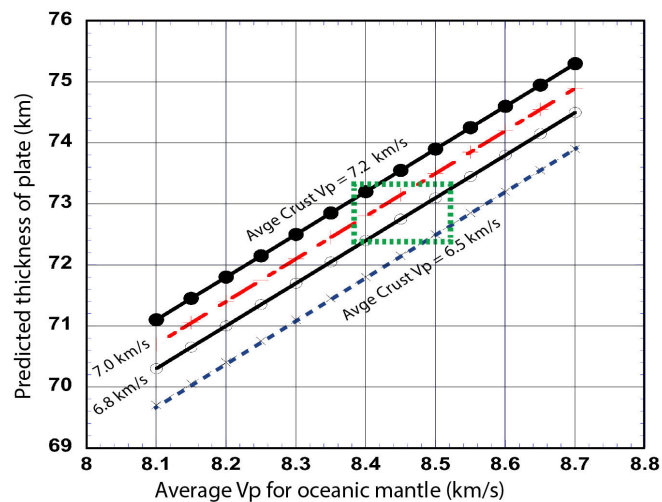
Extended Data Figure 4 | Stacking and model sensitivities. a, Stacking chart for image shown in Fig. 3a. Colour scale shows magnitude of the fold, orange stars are shot points. Numbers on the grid are metres north and east from the New Zealand Map Grid. **b,** Plots of seismic velocity versus depth[12] and of stacking velocity (the root-mean-squared average seismic velocity between the surface and a specific depth) versus depth used in the processing of the seismic section (Fig. 3). **c,** Plot of predicted plate thickness versus average v_P for the oceanic mantle; four plots are shown, each one derived using a different value of average crust v_P (labelled). Based on the position in travel time for the reflectors at zero offset on shot 11 (Extended Data Fig. 2c), we take the travel time in the oceanic crust to be 4 s and that in the oceanic mantle lid to be 14 s. The green dashed box represents the preferred range of solutions based on likely velocities for the oceanic crust and oceanic upper mantle[19]. Note that the uncertainty in the thickness of plate determination is $< \pm 1$ km.

a/

b/
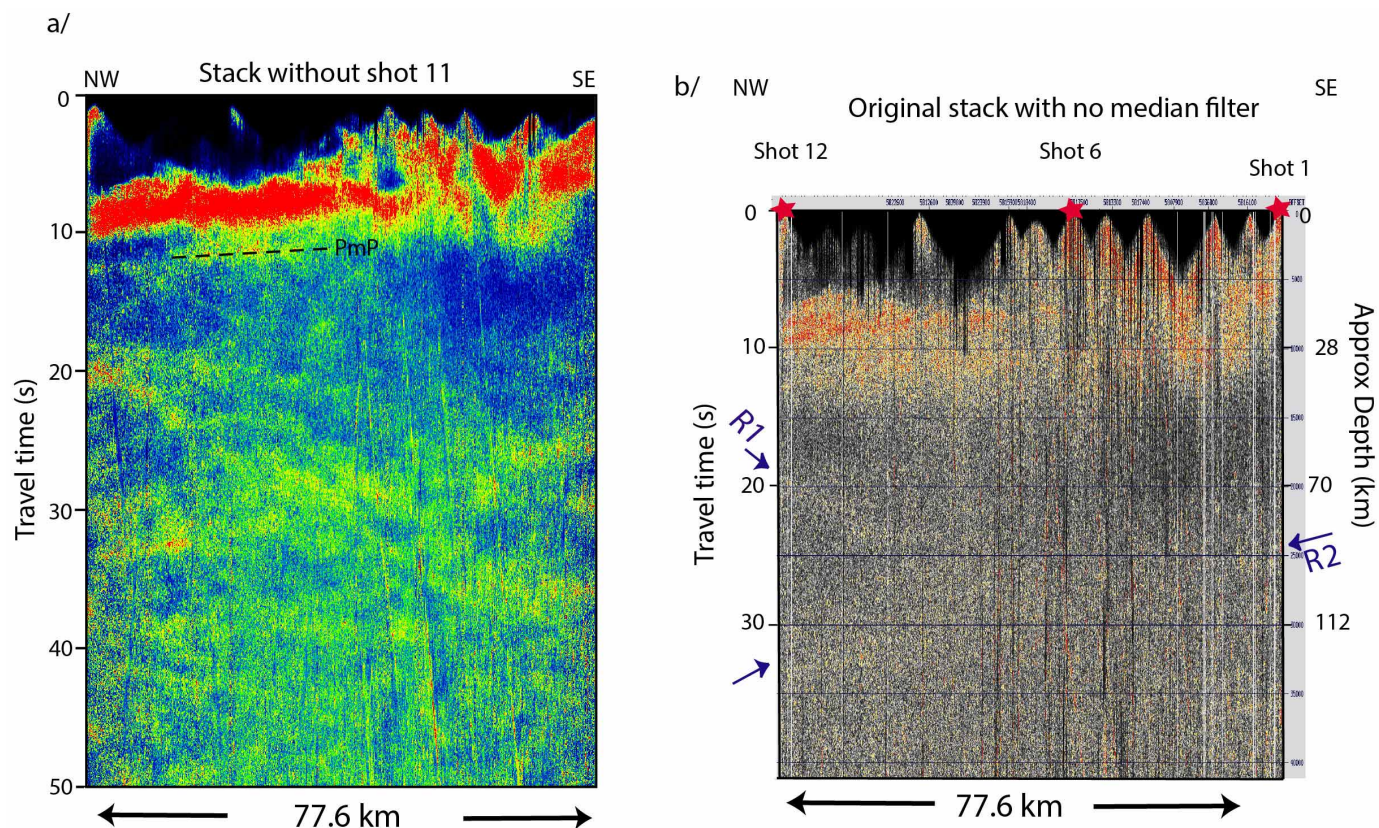
Extended Data Figure 5 | Stacking tests. a, Stack as for Fig. 3a but without shot 11. This is a check to ascertain how much the high-quality, higher-frequency shot 11 dominated the stack. We still see the main features in this stack, suggesting that the other shots are making a significant contribution. Dashed line shows our interpreted position for the Moho of the oceanic Pacific plate (P_mP reflection). **b**, A stack with no median filter applied.
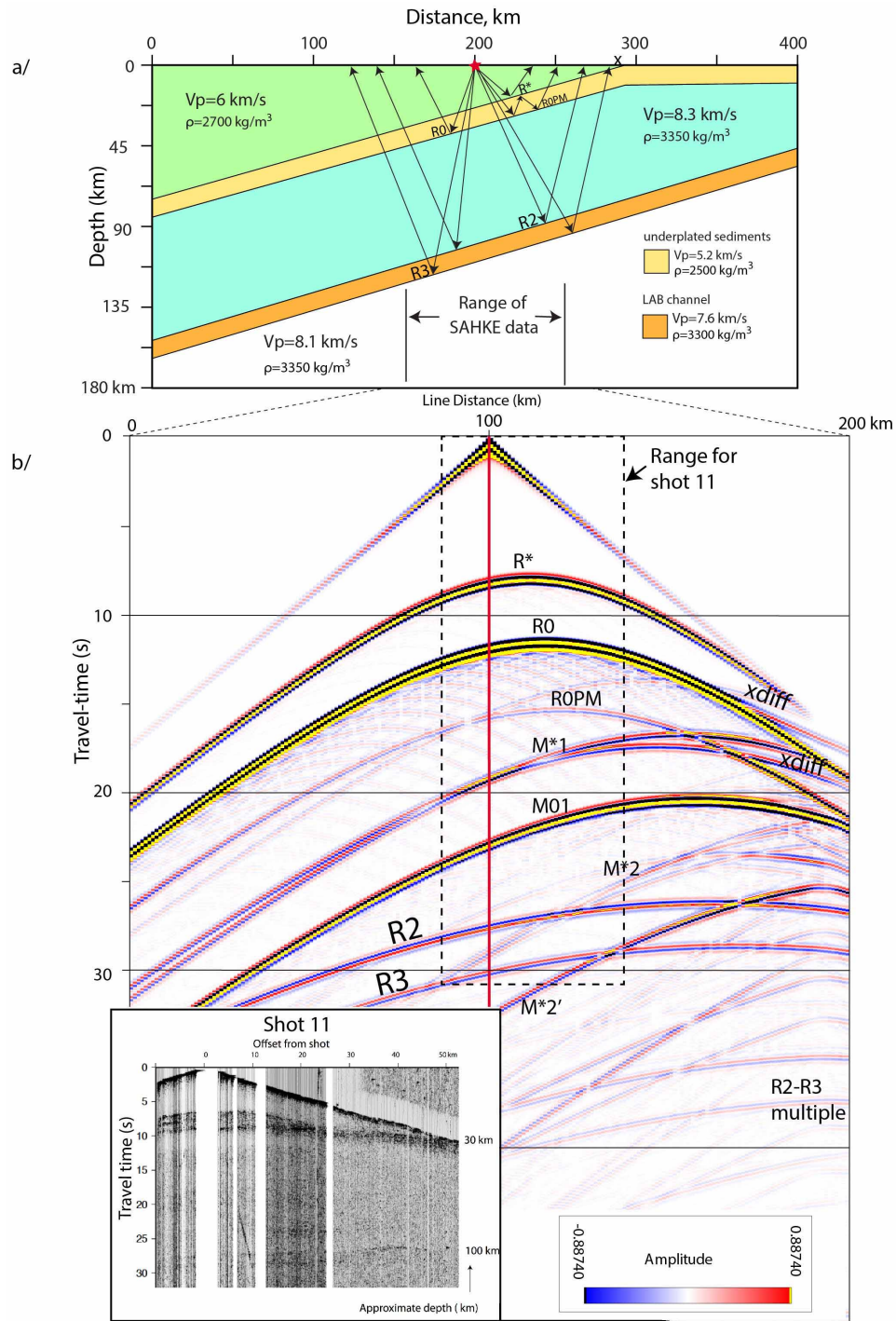
Extended Data Figure 6 | Wave-equation modelling for a dipping plate model. a, A model that simulates a highly simplified (oceanic crust not included) SAHKE structure for input into wave equation modelling. A layer of low-wave-speed sediments is included at the base of the crust, so we can examine how multiples from this channel interfere with the proposed R2 and R3 reflections at the base of the plate. The simulation is based on e3d[35] and is run with $v_S = 0$, so no S-wave reflections are created. **b**, Synthetic shot gather for shot 11 geometry based on the model. R* and R0 are respectively reflections from the top and the bottom of the sedimentary layer; R2 and R3 are reflections from the LAB channel. The M*1 and M01 events are interpreted as multiples of R* and R0, respectively, based on the lateral movement of the apex for the reflection hyperbola (see Methods). The second multiple appears to cross the R2 and R3 reflections, but is doing so at such a high angle it would be detected if present. Inset, the shot gather for shot 11 where no multiples can be seen at all. This is interpreted to mean multiples are not efficiently created where surface topography is rough, as it is for much of the SAHKE line. xdiff is diffraction from the pinch-out of the crustal layer marked as x in **a**. M*2 is interpreted as a second order multiple of R*.

a/

In-line and cross-line migration paths

- SAHKE Receivers
- In-line Migration
- X-line Migration

d/

SAHKE X-line Section

No obvious structure

b/

SAHKE In-line section from smoothed 0-phase records

Kirchhoff-sum Amplitude

-210677
210677

Top of plate structure

Base of plate reflections

e/

SAHKE X-line section from randomized smoothed records

c/

SAHKE In-line section randomized from smoothed 0-phase records

Kirchhoff-sum Amplitude

-210677
210677

f/

Extended Data Figure 7 | Pre-stack depth migration tests. a, The two azimuths used for 3D Kirchoff-sum, pre-stack depth migration[16] are in-line with the seismic line and at right angles to it (the cross-line). The velocity model used was an earlier version of that shown in Extended Data Fig. 5b with slightly lower velocities. The input shot gathers for 3D migration were the instantaneous amplitude-attribute traces, heavily smoothed by median filtering as in Fig. 3b, then bandpass filtered to produce an approximate zero-phase, low-frequency wavelet at each pre-stack reflection with strong amplitude. **b**, The in-line migration, showing coherency in the top 30 km with the dip of the top of the plate evident. Here constant travel-time diffraction arcs are created from each shot point. Where energy is enhanced we interpret there to be reflectors, and conversely if there is no enhancement, but just the diffraction arcs, we interpret this to be a lack of coherent structure. Coherent energy is seen at greater depths (90–100 km), which we ascribe to structure of the LAB. **c**, Here the data of each shot gather are randomized (jack-knife test[16]), and put through the same migration. Comparing the randomized with the correct data migration shows what are artefacts of the migration geometry, and what is signal. This confirms that images between 90 and 100 km depth seen in **b** are real. **d**, Result of migration in the cross-line direction; no coherent alignments are seen. **e**, Jack-knife test for the cross-line migration. Note that **e** looks similar to **d**, suggesting there is no alignment of structure in the cross-line direction, and that the reflections we are imaging are not side-swipe from out of plane structures. **f**, Schematic plot showing the relationship between the normalized thickness of the acoustic impedance transition zone, d/λ, and the relative (that is, normalized to the value for a vanishingly thin transition zone) reflection coefficient. Here d is the thickness of the zone, λ is the seismic wavelength, and the reflection coefficient is normalized to that for a impedance contrast of zero thickness[14]. Inset, schematic of the transition zone d between two layers of different velocities, v_1 and v_2.
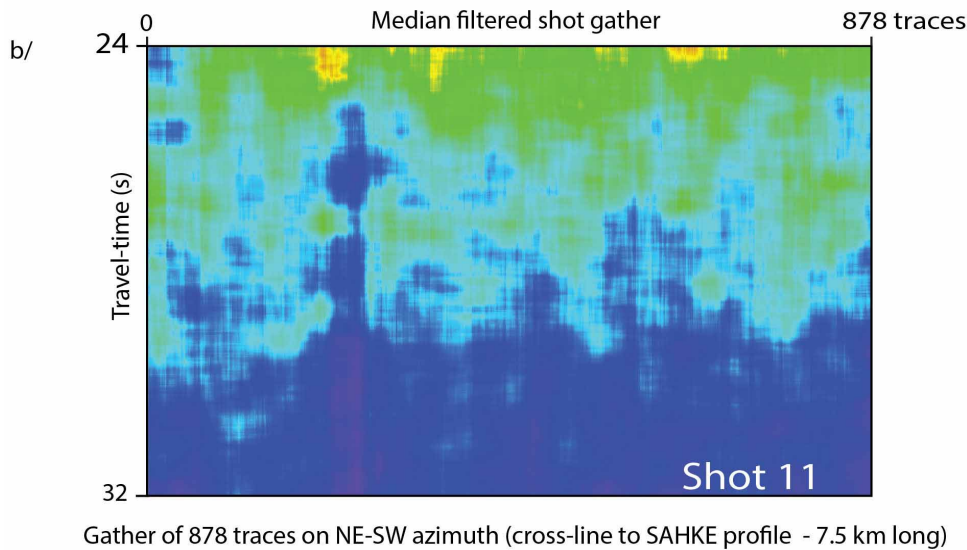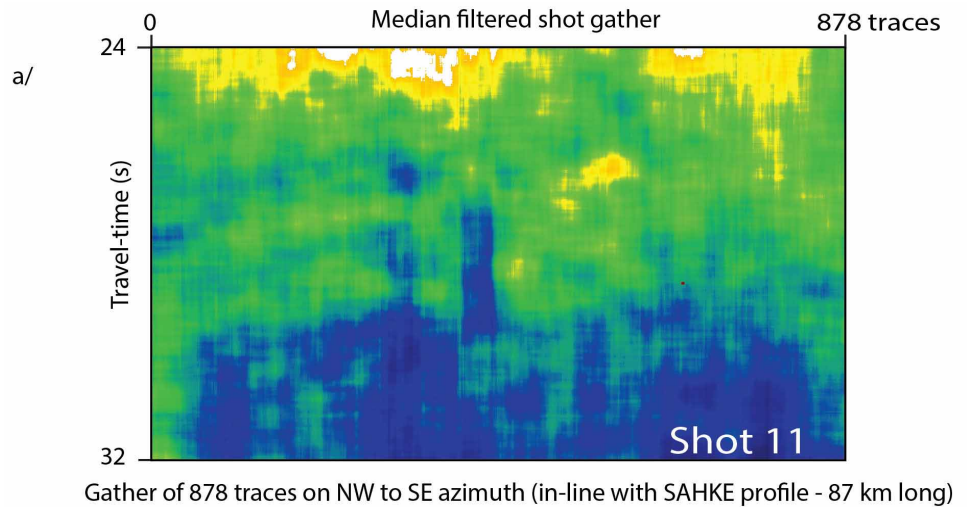
Extended Data Figure 8 | Gathers for shot 11 where the offsets are plotted in-line and cross-line. In **a**, the maximum offset from the shot is 70 km in a northwest to southeast azimuth (see Extended Data Fig. 7a). In **b**, the maximum offset of seismic receivers from shot 11, in a southwest to northeast azimuth, is 7.5 km. These are gathers of median filtered data (0.5 s by 41 traces), smoothed and passed through a 8–25 Hz bandpass filter, plotted in the 24–32 s range to bring out the R2 and R3 reflectors. Note the coherency in the in-line direction and lack of coherency in the cross-line gather. This further corroborates evidence from the pre-stack depth migration that the R2 and R3 reflections come from a surface below the northwest–southeast striking SAHKE04 line and are not side-swipe from out of plane reflections.
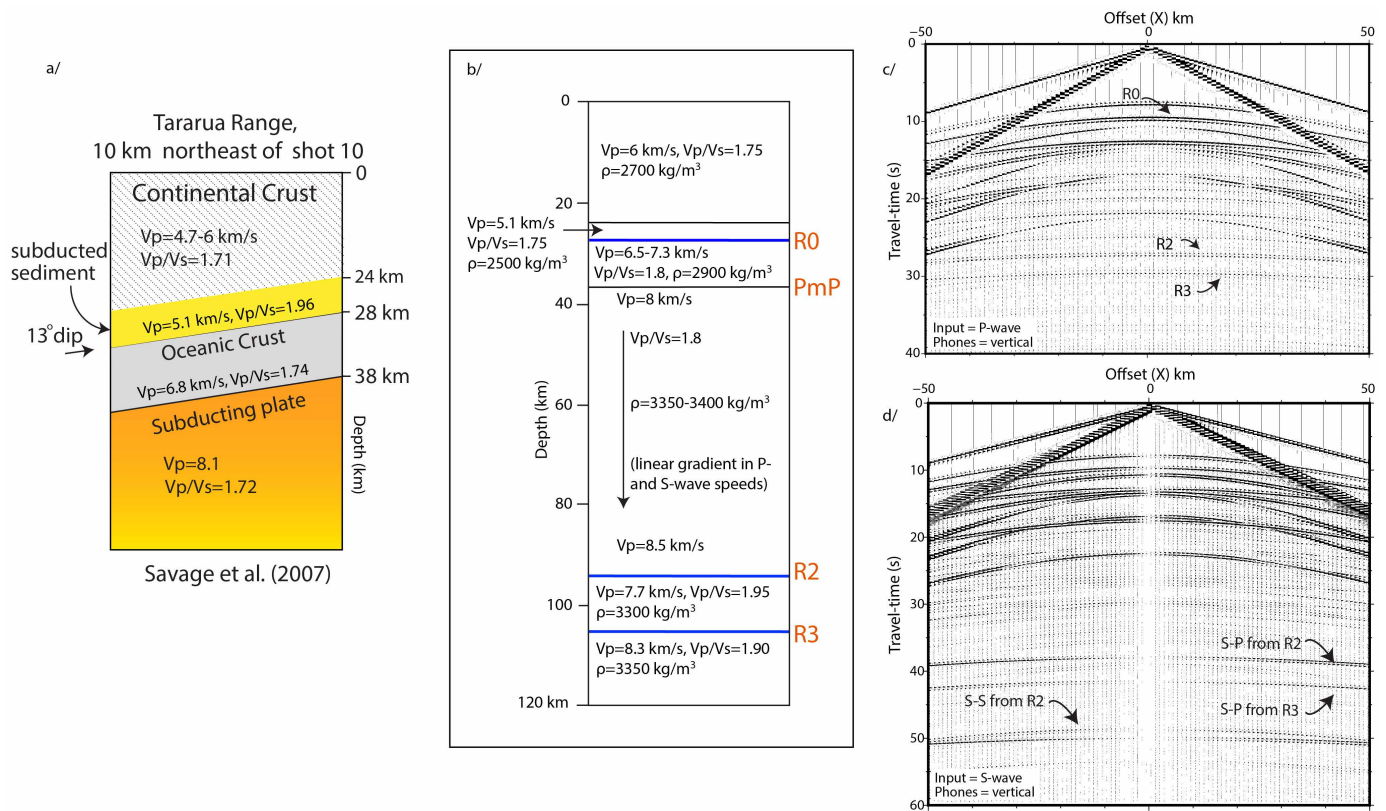
Extended Data Figure 9 | Models. a, Model of the structure and seismic velocities just north of the SAHKE line based on receiver functions[28]. Note the proposed subduction channel of accreted sediments, in line with more recent work[10], and the interpreted 13° dip, which is in the 12°–15° range proposed in this study. **b,** A simple horizontal-layered model which is used as input for the synthetic wave equation modelling using software e3d[35] shown in **c.** This is a simplified approximation model for our observations beneath SAHKE04, although we ignore dip. **c,** Wave equation modelling[35] based on input modelling shown in **b** for an input P wave and vertical geophones. No gain control is applied. Strong primary reflections and multiples from the top of the plate and oceanic Moho are shown. In reality, the surface multiples are not that strong because of scattering from topography. The R2 and R3 reflections can be seen as being weak events compared to R0. **d,** Wave equation modelling[35] based on input modelling shown in **b** for a input S wave and vertical geophones. Here the S-wave reflections are more prominent and S–P converted phases of R2 and R3 are predicted at about 38 and 41 s. Note that the S–P and S–S reflections have no energy at zero incidence angle and significant energy only for offsets >30 km. As we are only stacking data with maximum offsets of 50 km, we expect that some, but limited, S–P energy is contributing to the R4 phase of Fig. 3a.

LETTER

Recoded organisms engineered to depend on synthetic amino acids

Alexis J. Rovner[1,2], Adrian D. Haimovich[1,2]*, Spencer R. Katz[1,2]*, Zhe Li[1,2], Michael W. Grome[1,2], Brandon M. Gassaway[2,3], Miriam Amiram[1,2], Jaymin R. Patel[1,2], Ryan R. Gallagher[1,2], Jesse Rinehart[2,3] & Farren J. Isaacs[1,2]

Genetically modified organisms (GMOs) are increasingly used in research and industrial systems to produce high-value pharmaceuticals, fuels and chemicals[1]. Genetic isolation and intrinsic biocontainment would provide essential biosafety measures to secure these closed systems and enable safe applications of GMOs in open systems[2,3], which include bioremediation[4] and probiotics[5]. Although safeguards have been designed to control cell growth by essential gene regulation[6], inducible toxin switches[7] and engineered auxotrophies[8], these approaches are compromised by cross-feeding of essential metabolites, leaked expression of essential genes, or genetic mutations[9,10]. Here we describe the construction of a series of genomically recoded organisms (GROs)[11] whose growth is restricted by the expression of multiple essential genes that depend on exogenously supplied synthetic amino acids (sAAs). We introduced a _Methanocaldococcus jannaschii_ tRNA:aminoacyl-tRNA synthetase pair into the chromosome of a GRO derived from _Escherichia coli_ that lacks all TAG codons and release factor 1, endowing this organism with the orthogonal translational components to convert TAG into a dedicated sense codon for sAAs. Using multiplex automated genome engineering[12], we introduced in-frame TAG codons into 22 essential genes, linking their expression to the incorporation of synthetic phenylalanine-derived amino acids. Of the 60 sAA-dependent variants isolated, a notable strain harbouring three TAG codons in conserved functional residues[13] of MurG, DnaA and SerS and containing targeted tRNA deletions maintained robust growth and exhibited undetectable escape frequencies upon culturing ~10^{11} cells on solid media for 7 days or in liquid media for 20 days. This is a significant improvement over existing biocontainment approaches[2,3,6–10]. We constructed synthetic auxotrophs dependent on sAAs that were not rescued by cross-feeding in environmental growth assays. These auxotrophic GROs possess alternative genetic codes that impart genetic isolation by impeding horizontal gene transfer[11] and now depend on the use of synthetic biochemical building blocks, advancing orthogonal barriers between engineered organisms and the environment.

The advent of recombinant DNA technologies in the 1970s established genetic cloning methods[14], ushering in the era of biotechnology. Over the past decade, synthetic biology has fuelled the emergence of GMOs with increased sophistication as common and valued solutions in clinical, industrial and environmental settings[1,4,5], necessitating the development of safety and security measures first outlined in the 1975 Asilomar conference on recombinant DNA[15]. While guidelines for physical containment and safe use of organisms have been widely adopted, intrinsic biocontainment—biological barriers limiting the spread and survival of microorganisms in natural environments—remains a defining challenge. Existing biocontainment strategies employ natural auxotrophies or conditional suicide switches where top safeguards meet the 10^{-8} NIH standard (http://osp.od.nih.gov/office-biotechnology-activities/biosafety/nih-guidelines) for escape frequencies (that is, one escape mutant

per 10^8 cells), but can be compromised by metabolic cross-feeding or genetic mutation[9,10]. We hypothesized that engineering dependencies on synthetic biochemical building blocks would enhance existing containment strategies by establishing orthogonal barriers not feasible in organisms with a standard genetic code.

Our approach to engineering biocontainment used a GRO lacking all instances of the TAG codon and release factor 1 (terminates translation at UAA and UAG), eliminating termination of translation at UAG and endowing the organism with increased viral resistance, a common form of horizontal gene transfer (HGT). The TAG codon was then converted to a sense codon through the introduction of an orthogonal translation system (OTS) containing an aminoacyl-tRNA synthetase (aaRS):tRNA pair, permitting site-specific incorporation of sAAs into proteins without impairing cellular fitness[11]. Leveraging these unique properties of the GRO, we sought to reintroduce the TAG codon into essential genes to restrict growth to defined media containing sAAs. We also eliminated the use of multi-copy plasmids, which reduce viability and growth[16], impose biosynthetic burden, persist poorly in host cells over time[17], and increase the risk of acquiring genetic escape mutants[3], by manipulating native chromosomal essential genes and integrating the OTS into the genome. To engineer synthetic auxotrophies, we chose essential genes of varying expression levels (Methods), many of whose functions (for example, replication or translation) cannot be complemented by cross-feeding of metabolites. Genes dispersed throughout the genome were selected to prevent a single HGT event from compromising containment.

We pursued three strategies to engineer dependence on non-toxic, membrane-permeable, and well-characterized sAAs through the introduction of TAG codons into essential genes: (1) insertion at the amino terminus; (2) substitution of residues with computationally predicted tolerances[18]; and (3) substitution of conserved[13] residues at functional sites (Fig. 1a). We initially pursued the first two strategies in a GRO containing an OTS optimized for the sAA p-acetyl-L-phenylalanine (pAcF, α; see Methods for a detailed explanation of the nomenclature used). Using multiplex automated genome engineering (MAGE)[12], we targeted 155 codons for TAG incorporation via 4 pools of oligonucleotides (Supplementary Tables 1 and 2) in permissive media containing pAcF and L-arabinose (aaRS induction) (Fig. 1b). After replica plating on non-permissive media lacking pAcF and L-arabinose, we isolated eight pAcF auxotrophs with one strain containing two TAGs in essential genes (Fig. 1c and Supplementary Table 3). To determine whether our strategy was capable of creating synthetic auxotrophs dependent on other sAAs, MAGE was used to mutagenize annotated residues in the sAA binding pocket of the pAcF aaRS (Supplementary Table 4) to accommodate p-iodo-L-phenylalanine (pIF, β) or p-azido-L-phenylalanine (pAzF, γ) in two strains. After MAGE-based incorporation of TAGs and selections on permissive and non-permissive solid media, we obtained 8 pIF and 23 pAzF auxotrophs harbouring 1–4 TAGs at 30 distinct loci across 20 essential genes (Supplementary Tables 3 and 5). Together, these data

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA. [2]Systems Biology Institute, Yale University, West Haven, Connecticut 06516, USA. [3]Department of Cellular and Molecular Physiology, Yale University, New Haven, Connecticut 06520, USA.
*These authors contributed equally to this work.
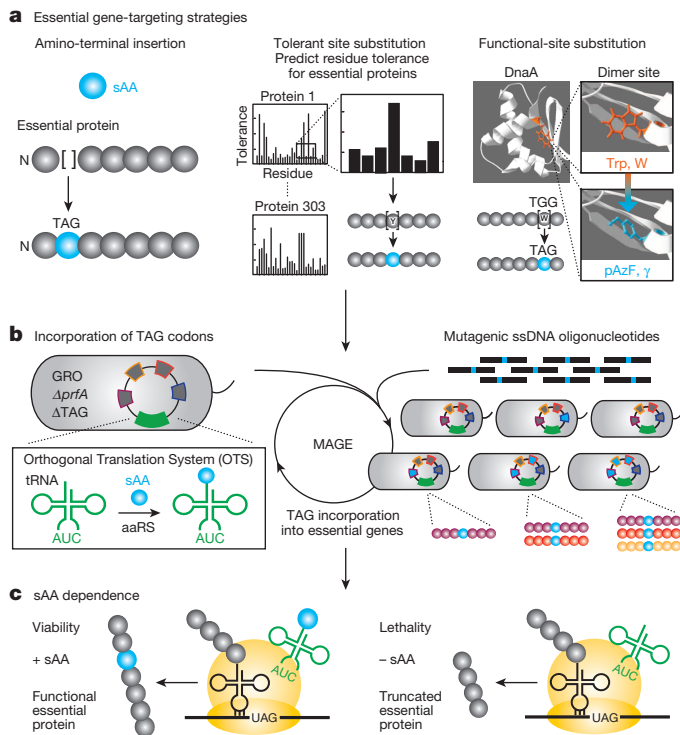
Figure 1 | Strategy used to engineer GROs to depend on sAAs for growth.
a, Approaches used to identify suitable loci within essential proteins for sAA (blue) incorporation. **b,** MAGE was used for site-specific incorporation of TAG codons into essential genes of a GRO lacking all natural TAG codons (ΔTAG) and release factor 1 ($\Delta prfA$), and containing an OTS (green) consisting of the *M. jannaschii* aaRS and cognate UAG-decoding tRNA. **c,** Synthetic auxotrophs that depend on sAAs for growth were isolated.

demonstrate the modularity of our approach and that synthetic auxotrophs can be engineered across many essential genes using multiple sAAs (Extended Data Fig. 1).

Measurements of doubling time in permissive media revealed minimal or no fitness impairment of synthetic auxotrophs relative to their non-contained ancestors with a genomically integrated OTS (Fig. 2a and Supplementary Table 6). To quantify the degree of containment, we measured the ratio of colony-forming units (c.f.u.) on non-permissive to permissive solid media and observed a range of escape frequencies spanning 10^{-3} to 10^{-7} (Fig. 2b). One notable strain DnaX.Y113α preserved the doubling time of its non-contained ancestor (Fig. 2a) while maintaining an escape frequency of 6.7×10^{-7} (Fig. 2b). We directly investigated pAcF incorporation in DnaX.Y113α using mass spectrometry and identified peptides containing pAcF at Y113 (Fig. 2c).

To investigate escape mechanisms of escape mutants derived from synthetic auxotrophs with one essential TAG codon, we performed targeted sequencing and observed transition mutations (A•T to G•C and G•C to A•T) commonly observed in mismatch-repair-deficient strains ($\Delta mutS$)[19]. All isolated DnaX.Y113α escape mutants incorporate tryptophan by mutation of the TAG codon to TGG. SecY.Y122α escape mutants incorporate glutamine by mutation of *glnV* to form a glutamine amber suppressor or mutation of the *secY.Y122* TAG codon to CAG (Supplementary Table 7). One of three SecY.Y122α escape mutants was wild type at the *secY.Y122* TAG codon and putative amber-suppressor loci[20], but whole-genome sequencing (Supplementary Table 8) revealed a Q54D missense mutation in *rpsD* (30S ribosomal subunit S4). This site is implicated in ribosome fidelity[21,22] and is the causal mutation leading to escape in this mutant (Extended Data Fig. 2).

These escape mechanisms informed two sets of experiments to engineer strains with lower escape frequencies. First, we sought to create synthetic auxotrophs with an increased numbers of TAGs (Fig. 2d) by combining TAGs from strains possessing the lowest escape frequencies
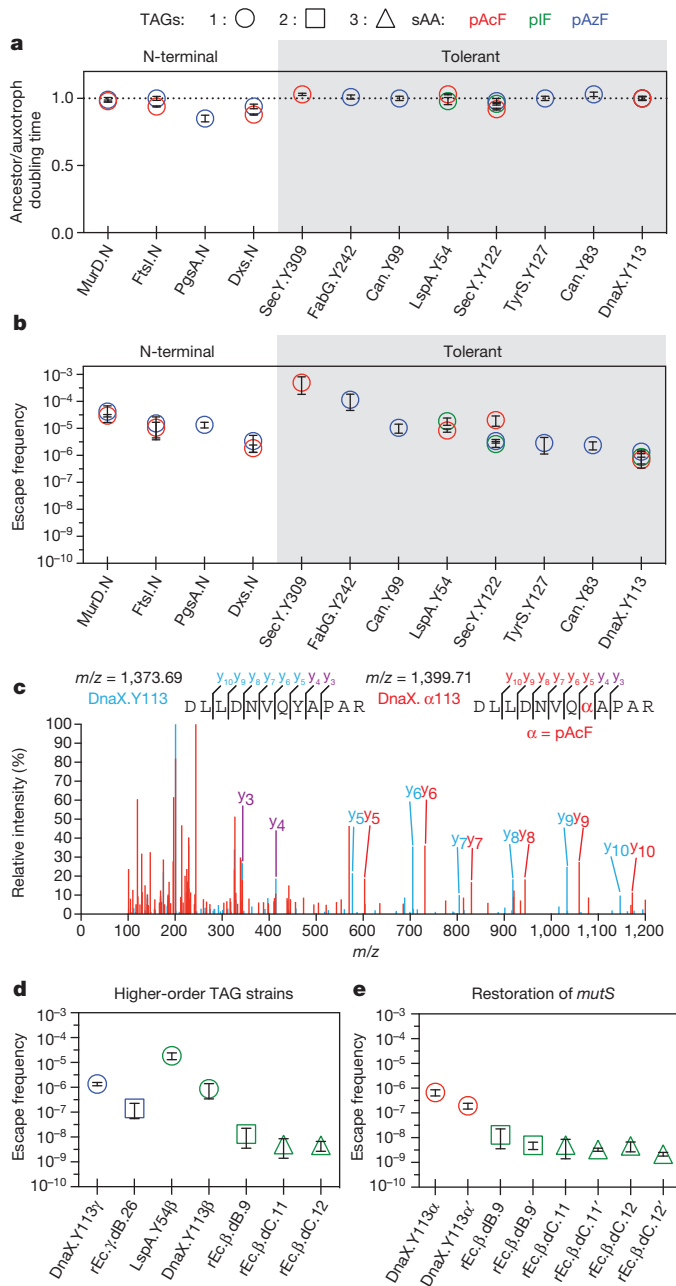
Figure 2 | Characterization of strains dependent on sAA incorporation in essential proteins. a, Doubling time ratios for the non-contained ancestor to synthetic auxotroph containing one TAG. **b,** Escape frequencies of strains from **a**. **c,** Superimposed MS/MS spectra for DnaX peptides from DnaX.Y113α (red) and the non-contained ancestor rEc.α (blue). Overlapping peaks are purple and a mass shift relative to rEc.α identifies Y113 as the pAcF incorporation site in DnaX.Y113α; see Methods. **d, e,** Escape frequencies for strains with multiple TAG codons (**d**) and/or functional mismatch repair (**e**) (prime, *mutS*$^{+}$). For all plots, average values of three technical replicates are plotted with error bars representing $\pm$ s.d. Reported results repeated at least three times in independent experiments.

(that is, *dnaX.Y113*, *lspA.Y54* and *secY.Y122*) into a single strain. In strains containing two TAGs, the escape frequency was reduced to 1.4×10^{-7} (rEc.γ.dB26) and 1.4×10^{-8} (rEc.β.dB9) (strain annotations are listed in Supplementary Table 6 and a complete description of our nomenclature can be found in the Methods). In strains containing three TAGs, escape frequencies were further reduced to 5.0×10^{-9} (rEc.β.dC.11) and 4.7×10^{-9} (rEc.β.dC.12). We used MAGE to assess quantitatively the effects of non-synonymous mutations at individual TAG codons in

strains incorporating pIF at SecY.Y122, DnaX.Y113, and LspA.Y54 by mutating the TAG site to sense codons for all 20 natural amino acids. Strains containing multiple TAGs were less likely to survive when one TAG was compromised (Extended Data Fig. 3). In a second set of experiments, we restored *mutS* (prime symbol (′) denotes a *mutS*[+] strain) and observed a decreased escape frequency in strains by 1.5- to 3.5-fold (Fig. 2e and Supplementary Table 6). Escape mutants derived from *mutS*[+] higher-order TAG strains exhibited impaired fitness with 1.14- to 1.28-fold greater doubling times than their contained ancestors. Whole-genome sequencing was performed on these escape mutants and revealed mutations of tyrosine tRNAs to form tyrosine amber (UAG) or ochre (UAA) suppressors (Supplementary Table 9).

To reduce escape frequencies below $\sim10^{-9}$ and eliminate rescue by natural amino acids, we pursued a third strategy to replace conserved and functional residues in essential proteins with sAAs (Fig. 1a). Using the Conserved Domain Database[13], we searched all essential proteins for tyrosine, tryptophan and phenylalanine residues involved in protein–protein interactions (for example, dimerization) or located within active sites to identify candidates suitable for replacement with phenylalanine-derived sAAs. After targeted insertion of TAG codons using MAGE, we isolated four synthetic auxotrophs with pAzF incorporated at GlyQ.Y226 (glycyl-tRNA synthetase α subunit, dimer interface), Lnt.Y388 (apolipoprotein *N*-acyltransferase, active site), MurG.F243 (*N*-acetylglucosaminyl transferase, active site), and DnaA.W6 (chromosomal replication initiator protein, oligomerization site[23]) in strains with minor fitness impairments (Fig. 3a) and escape frequencies spanning 10^{-5} to 10^{-7} (Fig. 3b). Identical experiments to incorporate pAcF and pIF failed to generate synthetic auxotrophs, suggesting that the targeted residues are recalcitrant
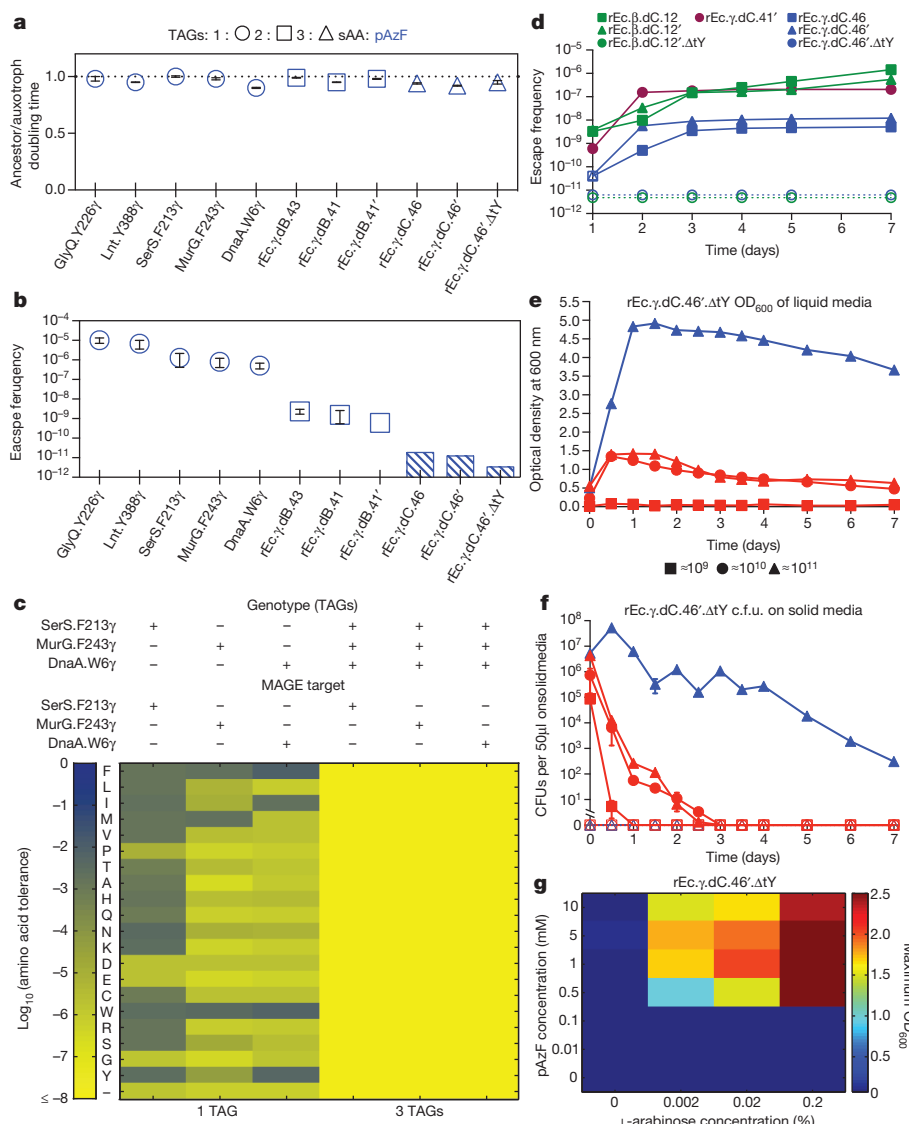
Figure 3 | Characterization of strains dependent on sAA incorporation at active and dimerization sites in essential proteins. a, Doubling time ratios for the non-contained ancestor to pAzF auxotroph with one or more TAGs at functional loci calculated from growth in 5 mM pAzF and 0.2% L-arabinose. **b**, Escape frequencies of strains in **a**; bars represent escape frequencies below the detection limit; average escape frequencies are plotted. **c**, Representative assay surveying tolerance of TAG loci to 20 amino acids in different synthetic auxotrophs and expressed as log₁₀ of total cell survival. A + symbol indicates a TAG codon at the locus in the background strain; – indicates the wild-type codon; blue and yellow indicate high and low tolerance to substitution, respectively; see Methods. **d**, Representative escape assay monitoring escape frequencies up to 7 days after plating on solid non-permissive media; hollow symbols/dashed lines, no observed escape mutants; see Methods. **e**, Temporal monitoring of permissive (P, blue) and non-permissive (NP, red) cultures inoculated with $\sim10^9$, 10^{10} or 10^{11} cells of rEc.γ.dC.46′.ΔtY by OD₆₀₀. **f**, Associated c.f.u. from **e** as sampled on permissive (solid lines) or non-permissive (dashed lines) solid media; c.f.u. were never observed on non-permissive solid media; hollow data points indicate no observed c.f.u. **g**, Maximum OD₆₀₀ values during growth in LB media across a concentration gradient of pAzF and L-arabinose. For all plots, average values of three technical replicates are plotted with error bars representing ±s.d. Reported results repeated at least three times in independent experiments.

to replacement by pAcF and pIF. Since pAzF was able to replace conserved and functional tyrosine, phenylalanine and tryptophan residues across several essential proteins, we hypothesized that engineering strains to contain higher-order TAG combinations would limit escape by mutations that cause incorporation of natural amino acids at multiple TAG codons. Escape frequencies of 1.6×10^{-9} and 2.3×10^{-9} were observed for strains containing two TAG codons: rEc.γ.dB.41 (DnaA.W6 and MurG.F243) and rEc.γ.dB.43 (DnaA.W6 and SerS.F213), respectively. Upon restoring *mutS*, the escape frequency of rEc.γ.dB.41′ fell to 6.0×10^{-10} (Fig. 3b). Merging all three sites into one strain (rEc.γ.dC.46) and its *mutS*$^{+}$ derivative (rEc.γ.dC.46′) led to escape frequencies of $<7.9 \times 10^{-11}$ and $<4.4 \times 10^{-11}$ (below the detection limit of our plate-based assay), respectively (Supplementary Table 6).

Temporal monitoring of rEc.γ.dC.46′ revealed the emergence of growth-impaired escape mutants 2 days post-plating on non-permissive solid media (Fig. 3d). Sequencing of escape mutants derived from strains rEc.γ.dC.41′ and rEc.γ.dC.46′ revealed amber-suppressor-forming mutations at one of three tyrosine tRNAs (*tyrT*, *tyrV*, *tyrU*) with growth impairments spanning 1.61- to 2.10-fold increases in doubling time relative to contained ancestors (Supplementary Tables 9 and 10). Given that *E. coli* contains three tyrosine tRNAs, we hypothesized that deletion of *tyrT* and *tyrV*[24] would prevent acquisition of amber-suppressor-forming mutations at *tyrU*, as preservation of this single remaining copy of tRNA^{Tyr} would be required to maintain fidelity of tyrosine incorporation during protein synthesis. We used λ-Red recombination to delete *tyrT* and *tyrV* in rEc.β.dC.12′, rEc.β.dC.12′.E7 (escape mutant of rEc.β.dC.12′), and rEc.γ.dC.46′ with a chloramphenicol resistance gene. Deletion of *tyrT* and *tyrV* restored containment of the escape mutant, establishing the causal escape mechanism (Extended Data Fig. 4). Moreover, *tyrT/V* deletions in rEc.β.dC.12′.ΔtY and rEc.γ.dC.46′.ΔtY decreased escape frequencies below detectable levels ($<4.9 \times 10^{-12}$ and $<6.3 \times 10^{-12}$, respectively) over the 7-day observation period (Fig. 3d and Supplementary Table 11).

To challenge strains rEc.β.dC.12′.ΔtY and rEc.γ.dC.46′.ΔtY with natural amino acids and mimic a potential HGT event, we introduced constructs containing phenylalanine or tryptophan amber-suppressor tRNAs. While growth of suppressor-containing strains was equivalent to the cognate-contained ancestor in permissive liquid media, severely impaired growth or no growth was observed in non-permissive media (Extended Data Fig. 5). Such findings are further supported in experiments where a large ($\sim 10^{11}$) inoculum of cells challenged on solid or in liquid (see below) non-permissive media do not yield escape mutants, providing ample opportunity for natural formation of a phenylalanine amber suppressor via mutation of one of two native copies of tRNA^{Phe}. These data support our hypothesis that synthetic auxotrophs containing higher-order TAG combinations depend on the sAA and limit growth from natural amino acids.

To interrogate the long-term stability of synthetic auxotrophs where escape mutant formation is not limited by a colony growth environment, temporal monitoring of rEc.γ.dC.46′.ΔtY was performed on large cell populations in liquid culture (1 litre of Luria–Bertani (LB) media) for 7 days with frequent OD$_{600}$ measurements to track cell growth (Fig. 3e). Inoculation of $\sim 10^{11}$ cells in permissive media led to a confluent culture of contained cells within 24 h. Inoculation of $\sim 10^{9}$, $\sim 10^{10}$ and $\sim 10^{11}$ cells into non-permissive media revealed transient growth, which we propose is due to residual pAzF and L-arabinose from large inoculums, followed by a sustained decrease in cell density and growth termination. Cell survival and escape from liquid cultures was monitored by quantifying c.f.u. on permissive and non-permissive solid media, respectively (Fig. 3f). Plating on permissive solid media revealed a drop in c.f.u. to below the limit of detection within 1 day from the non-permissive flask inoculated with $\sim 10^{9}$ cells and 3 days from non-permissive flasks inoculated with $\sim 10^{10}$ or $\sim 10^{11}$ cells (Fig. 3f). No c.f.u. were observed from any culture plated on non-permissive solid media. To confirm the absence of a single escape mutant following an extended 20-day growth period (Extended Data Fig. 6), the non-permissive and permissive cultures

inoculated with $\sim 10^{11}$ cells were plated across 30 non-permissive plates. Escape mutants were not observed and escape frequencies remained below the detection limit after 7 days, which is comparable to the solid media results. These results demonstrate that rEc.γ.dC.46′.ΔtY depends on pAzF, maintains long-term stability of biocontainment in permissive liquid media and exhibits termination of growth in non-permissive media.

To further investigate the dependency on sAAs, liquid growth profiles were collected for synthetic auxotrophs across sAA and L-arabinose concentration gradients. Growth of rEc.γ.dC.46′.ΔtY was not observed below 0.002% L-arabinose and 0.5 mM pAzF (Fig. 3g and Extended Data Fig. 7). Growth increased in a dose-dependent manner with increasing concentrations of pAzF and L-arabinose, where 5 mM pAzF and 0.2% L-arabinose was optimal for fitness (that is, maximum OD$_{600}$ and minimum doubling time). In an equivalent experiment with rEc.β.dC.12′.ΔtY, 1 mM pIF and 0.2% L-arabinose was optimal for fitness (Extended Data Fig. 8). Since growth was not observed in media lacking either L-arabinose or the sAA, these data further support the dependency of synthetic auxotrophs on sAAs.

To determine whether a synthetic auxotroph could be rescued by metabolic cross-feeding, we evaluated the viability of strains on diverse media types. We grew wild-type MG1655 *Escherichia coli*, a biotin auxotroph (EcNR2[12]), a non-contained GRO (rEc.γ), and the pAzF synthetic auxotroph (rEc.γ.dC.46′) on solid media containing both pAzF/L-arabinose and biotin, either pAzF/L-arabinose or biotin, and on plates lacking small molecules (Fig. 4). Despite biotin auxotrophy, growth of EcNR2 on rich defined media without biotin was rescued in close proximity to wild-type *E. coli*, suggesting cross-feeding of essential metabolites (Extended Data Fig. 9). Blood agar and soil extracts without biotin or pAzF/L-arabinose supplementation supported growth of all strains except the synthetic auxotroph, which only grew on media supplemented with pAzF and L-arabinose. These data suggest that synthetic auxotrophies could lead to a more viable containment strategy for clinical (for example, blood) and environmental (for example, soil) settings, where metabolic auxotrophies can be overcome by proximal, metabolically competent strains.

Synthetic auxotrophs utilize unnatural biochemical building blocks necessary for essential proteins with activities that cannot be complemented by naturally occurring small molecules. We have previously shown that genomic recoding interferes with HGT from viruses[11], and have now extended orthogonal barriers by engineering two synthetic auxotrophs using two unique sAAs that exhibit escape frequencies below our detection limit ($<6.3 \times 10^{-12}$). These synthetic auxotrophs possess three essential TAGs at loci dispersed throughout the genome (0.84, 0.86 and 2.9 megabases apart), thereby limiting the likelihood that a single HGT event could compromise containment (Extended Data Fig. 1). These orthogonal barriers can be expanded further by incorporation of additional TAG sense codons across more than three essential
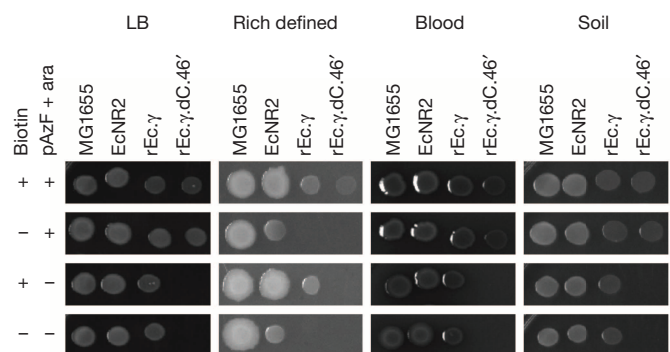

Figure 4 | Investigating the viability of synthetic auxotrophs on diverse media types. Rescue by cross-feeding shown through spotting on diverse media types with or without pAzF/L-arabinose (pAzF + ara) and biotin supplementation; EcNR2, rEc.γ, and rEc.γ.dC.46′ are auxotrophic for biotin (*ΔbioA/B*) and rEc.γ.dC.46′ is also a pAzF auxotroph.

genes, but will probably require concurrent advances in OTS performance to maintain fitness and viability (for example, enhanced activity and specificity of aaRS:tRNA pairs[25]). Our modular approach to biocontainment limits growth to synthetic environments containing unnatural biochemical building blocks with diverse chemistries. We anticipate that further genome recoding efforts[11,26,27] will enable auxotrophies for multiple sAAs that could be enhanced by other synthetic components including unnatural nucleotides and extended genetic alphabets[28–30]. Orthogonal biological systems employing multi-level containment mechanisms are uniquely suited to provide safe GMOs for clinical, environmental and industrial applications[1].

Despite the breadth of genomic diversity found in nature, all species utilize the same biochemical foundation to sustain life. The semantic architecture of the GRO employs orthogonal translational components, establishing the basis for a synthetic molecular language that relieves limitations on natural biological functions by depending on the incorporation of sAAs with exotic chemistries. This work sets the stage for future experiments to probe the optimality of the natural genetic code and to explore the plasticity of proteins and whole organisms capable of sampling new evolutionary landscapes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* **157,** 151–161 (2014).
2. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered biology. *FEBS Lett.* **586,** 2199–2206 (2012).
3. Moe-Behrens, G. H., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4,** 5 (2013).
4. Pieper, D. H. & Reineke, W. Engineering bacteria for bioremediation. *Curr. Opin. Biotechnol.* **11,** 262–270 (2000).
5. Steidler, L. Genetically engineered probiotics. *Best Pract. Res. Clin. Gastroenterol.* **17,** 861–876 (2003).
6. Kong, W. *et al.* Regulated programmed lysis of recombinant *Salmonella* in host tissues to release protective antigens and confer biological containment. *Proc. Natl Acad. Sci. USA* **105,** 9361–9366 (2008).
7. Szafranski, P. *et al.* A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proc. Natl Acad. Sci. USA* **94,** 1059–1063 (1997).
8. Steidler, L. *et al.* Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nature Biotechnol.* **21,** 785–789 (2003).
9. Ronchel, M. C. & Ramos, J. L. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl. Environ. Microbiol.* **67,** 2649–2656 (2001).
10. Jensen, L. B., Ramos, J. L., Kaneva, Z. & Molin, S. A substrate-dependent biological containment system for *Pseudomonas putida* based on the *Escherichia coli gef* gene. *Appl. Environ. Microbiol.* **59,** 3713–3717 (1993).
11. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–360 (2013).
12. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
13. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33,** D192–D196 (2005).
14. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* **70,** 3240–3244 (1973).
15. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O. & Singer, M. F. Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc. Natl Acad. Sci. USA* **72,** 1981–1984 (1975).
16. Betenbaugh, M. J., Beaty, C. & Dhurjati, P. Effects of plasmid amplification and recombinant gene expression on the growth kinetics of recombinant *E. coli*. *Biotechnol. Bioeng.* **33,** 1425–1436 (1989).
17. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer *in situ*: a critical review. *Nature Rev. Microbiol.* **3,** 700–710 (2005).
18. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4,** 1073–1081 (2009).
19. Schaaper, R. M. & Dunn, R. L. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc. Natl Acad. Sci. USA* **84,** 6220–6224 (1987).
20. Eggertsson, G. & Soll, D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol. Rev.* **52,** 354–374 (1988).
21. Maisnier-Patin, S., Berg, O. G., Liljas, L. & Andersson, D. I. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol. Microbiol.* **46,** 355–366 (2002).
22. Holberger, L. E. & Hayes, C. S. Ribosomal protein S12 and aminoglycoside antibiotics modulate A-site mRNA cleavage and transfer-messenger RNA activity in *Escherichia coli*. *J. Biol. Chem.* **284,** 32188–32200 (2009).
23. Abe, Y. *et al.* Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* **282,** 17816–17827 (2007).
24. Bösl, M. & Kersten, H. A novel RNA product of the tyrT operon of *Escherichia coli*. *Nucleic Acids Res.* **19,** 5863–5870 (1991).
25. O'Donoghue, P., Ling, J., Wang, Y. S. & Soll, D. Upgrading protein synthesis for synthetic biology. *Nature Chem. Biol.* **9,** 594–598 (2013).
26. Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333,** 348–353 (2011).
27. Lajoie, M. J. *et al.* Probing the limits of genetic recoding in essential genes. *Science* **342,** 361–363 (2013).
28. Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* **356,** 537–539 (1992).
29. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336,** 341–344 (2012).
30. Hammerling, M. J. *et al.* Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nature Chem. Biol.* **10,** 178–180 (2014).

Author Contributions A.J.R. and F.J.I. conceived the study, designed experiments and wrote the paper with assistance from A.D.H. and S.R.K.; A.J.R. conducted experiments with assistance from S.R.K., A.D.H., Z.L., M.W.G., M.A., J.R.P. and R.R.G.; B.M.G. and J.R. conducted mass spectrometry. All authors commented on the paper and F.J.I. supervised all aspects of the study.

Author Information Genome sequences have been deposited in GenBank under the accession numbers CP010455 (rEc.γ.dC.46) and CP010456 (rEc.β.dC.12). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.J.I. (farren.isaacs@yale.edu).

METHODS

Reagents. Oligonucleotide synthesis was performed by Integrated DNA Technologies (IDT) and Keck Foundation Biotechnology Resource Laboratory at Yale University (Supplementary Table 1). Unless otherwise stated, all cultures were grown in LB media. The following selective agents and inducers were used at the specified concentrations: ampicillin (amp, 50 μg ml^{-1}), carbenicillin (carb, 50 μg ml^{-1}), zeocin (zeo, 10 μg ml^{-1}), spectinomycin (spec, 95 μg ml^{-1}) and sodium dodecyl sulphate (SDS, 0.005% w/v), isopropyl-β-D-1-thiogalactopyranoside (IPTG, 100 μM), 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal, 40 μg ml^{-1}), and L-arabinose (ara, 0.2% w/v unless otherwise indicated). sAAs were used at 1 mM unless otherwise indicated and purchased from PepTech (pAcF, AL624-2), BaChem (pIF, F-3075.0005) and Chem-Impex International (pAzF, 03376).

Plasmids. All tRNAs used to assess tolerance for tryptophan and phenylalanine at TAG codons were contained within the pTech plasmid backbone and driven by the *lpp* promoter[31]. Isothermal assembly[32] was used to replace the chloramphenicol acetyltransferase (*cat*) gene with the *sh ble* gene for resistance to zeocin.

The *supU* amber suppressor tRNA[33] was used to assess tolerance for tryptophan and a phenylalanine amber suppressor[31] was used to assess tolerance for phenylalanine. pTech-*supU* was provided by the laboratory of D. Söll and *supPhe* was synthesized by IDT and isothermally assembled into the pTech plasmid backbone to obtain pTech-*supPhe*.

Conversion of aminoacyl-tRNA synthetase specificity. The pAcF OTS was integrated into the genome of the GRO linked to a counter-selectable gene *tolC*. Co-selection multiplex automated genome engineering (CoS-MAGE[34]) was used as described previously to introduce annotated mutations[35] to the sAA binding pocket of the aaRS for specificity towards pAzF or pIF (Supplementary Table 4). Sanger sequencing was used to verify these mutations. Conversion of sAA-specificity was assessed in sequence-verified clones upon growth in the presence of sAA incorporation and episomally-expressed GFP containing an in-frame TAG codon at residue 151 within the protein product. OTS-mediated suppression of this codon with the sAA (that is, pAzF, pIF) generated a full-length fluorescent product, indicating that sAA incorporation had occurred and specificity was achieved.

TAG codon incorporation into essential genes. We applied three unique strategies to identify permissive sites in essential genes for TAG codon incorporation (Fig. 1). In our first strategy, a subset of essential genes[36] were chosen for the incorporation of one or more TAG codons immediately after the start codon to encode a sAA at the amino terminus. To explore a diverse library of incorporation targets within the *E. coli* proteome, in our second strategy we applied the sorting intolerant from tolerant (SIFT) algorithm[18] (downloaded on the Yale Biomedical High Performance Computing Cluster, http://sift.jcvi.org/) to the entire panel of essential *E. coli* proteins[36]. SIFT is an algorithm that uses sequence homology to predict the tolerance of amino acid substitutions at different indices. In our workflow, genes were first split into three categories on the basis of wild-type expression level[37], and a further four subgroups by genomic location with the goal of targeting essential genes dispersed throughout the *E. coli* chromosome. Next, genes shown to be essential by multiple studies[36] were passed through SIFT. For each essential gene, two high, medium, and low tolerance sites were targeted for TAG incorporation by MAGE. By this approach, we were able to sample diverse residue types in proteins with varying wild-type expression levels.

In our third strategy, using the conserved domain database[13], we searched within all annotated essential proteins for tyrosine, tryptophan and phenylalanine residues predicted to participate in essential enzymatic reactions or protein–protein interactions (e.g., dimerization). To minimize the probability that the added functionality of the sAA would perturb protein function, we targeted sites that were observed to occur as tyrosine or tryptophan in different homologues.

GROs containing an OTS integrated into the chromosome were grown to mid-log phase in liquid permissive LB media and four cycles of MAGE were performed per pool of mutagenic oligonucleotides (oligonucleotide concentration ≤ 15 μM) as described previously[12,38]. To isolate synthetic auxotrophs, mutagenized cultures were plated on solid media and replica plated onto non-permissive media. To identify TAG incorporation loci, multiplex allele-specific colony (MASC) PCR was used to interrogate pools of up to eleven targeted loci as previously described[38], followed by verification using Sanger sequencing.

Genotyping. Sanger sequencing was performed by the Keck DNA Sequencing Facility at Yale University or by GENEWIZ, Inc. Genomic DNA for whole genome sequencing was prepared using a Qiagen Genomic DNA purification kit. Illumina libraries were prepared by the Yale Center for Genomic Analysis or the Dana Farber Cancer Institute. Illumina HiSeq or MiSeq sequencing systems were used for whole genome sequencing to generate 50- or 150-base-pair (bp) paired-end reads, respectively.

Whole-genome sequencing was used to analyse three escape mutants per background. In all cases, the direct ancestor to the escape mutant was also analysed. SNPs in escape mutants were identified relative to the reference genome *E. coli* C321.ΔA

(CP006698.1, GI:54981157) using a previously described[11] software pipeline. SNPs listed in Supplementary Table 8 and Supplementary Table 10 were called by Freebayes in escape mutants.

Strains. All GROs used in this study are derived from C321.ΔA (CP006698.1, GI:54981157)[38] which lacks all TAG codons and release factor 1. This strain is derived from strain EcNR2 (*ΔmutS:catΔ*(*ybhB-bioAB*):[*cI857Δ*(*cro-ea59*):*tetR-bla*]), modified from *E. coli* K-12 substr. MG1655. In all synthetic auxotrophs, the *M. jannaschii*-derived OTS was genomically integrated into the GRO fused to the counter-selectable gene *tolC*. The OTS consists of an L-arabinose-inducible aaRS driven by the *araBAD* promoter, and a constitutively expressed cognate amber-decoding tRNA driven by the *proK* promoter. All genome modifications that required incorporation of dsDNA (for example, modifications to the *mutS* gene or incorporation of antibiotic selectable markers) were performed via λ-Red recombination[38].

Nomenclature of genomically recoded organisms and synthetic auxotrophs. To succinctly name strains, we have introduced a new one-letter amino acid code for sAAs using Greek lettering (pAcF = α, pIF = β, and pAzF = γ). Non-contained GROs lacking essential TAG codons are named according to the one letter sAA code for the specific OTS present in the organism. For example, a ΔTAG GRO with a genomically integrated pAcF OTS is rEc.α.

Biocontained GROs containing essential TAG codons are named according to two conventions based on the number of essential TAG codons in the auxotroph: (1) Strains with one essential TAG are named by the essential protein containing the sAA and the position and identity of the residue substituted therein (for example, a strain containing pAcF at residue 113 in DnaX is DnaX.Y113α); (2) Strains containing more than one essential TAG are named using the one letter sAA code for which the organism is auxotrophic. This is followed by a dependency code, d, indicating the presence of two (dB), three (dC) or four (dD) essential TAG codons, and then by a TAG combination number that uniquely identifies the specific combination of TAGs in the strain. Combinations are numbered from one through 46 and are listed in Supplementary Table 6.

Mismatch repair. The presence of a prime symbol (′) following the TAG combination number indicates that *mutS* has been restored at its native locus, imparting functional mismatch repair to the organism.

tRNA redundancy. Following the TAG combination number, Δt indicates the amino acid for which tRNA redundancy has been eliminated and is followed by the relevant amino acid (for example, a strain in which two of three total tyrosine tRNAs were deleted is ΔtY).

Escape mutant identity. At least three escape mutants were characterized per strain background that permitted an escape mutant. An escape mutant is designated by a number following the letter 'E' (for example, E1).

The summary of synthetic auxotrophs generated in this study illustrated in Extended Data Fig. 1 was constructed using the Circos[39] software.

Strains were grown at 34 °C in flat-bottomed 96-well plates containing 150 μL of LB media permissive for sAA incorporation, unless otherwise indicated. Strains were washed twice with sterile dH$_2$O before assessing growth in non-permissive media. Kinetic growth (OD$_{600}$) was monitored on a BioTek plate reader at ten-minute intervals in triplicate. Raw OD$_{600}$ data from the plate reader were normalized to standard absorbance (OD$_{600}$ at 1 cm path length) values using an empirically derived calibration curve ($Y = 1.9704x - 0.1183$, where y = OD$_{600}$ at 1 cm path length and x = OD$_{600}$ from plate reader; $R^2 = 0.998$). DTs were calculated in MATLAB using custom code. Reported values are the average between three technical replicates where error bars represent ± s.d. All reported results repeated at least three times in independent experiments. Maximum OD$_{600}$ values were obtained after 24 h of growth and represent the average of three technical replicates. Reported results repeated at least three times in independent experiments.

Mass spectrometry. Histidine-tagged proteins were purified on NiNTA resin (Qiagen). Resolution of purity was assessed via SDS–PAGE. In-gel digestion was performed similarly to previously described methods[40]. Proteins were stained and imaged within the gel using Coomassie blue (R-250). A band corresponding to the molecular weight of DnaX was excised. Gel slices were processed into 1-mm cubes, washed in 1:1 (v/v) 50% CH$_3$CN/50 mM NH$_4$HCO$_3$, and then washed in 1:1 (v/v) 50% CH$_3$CN/10 mM NH$_4$CO$_3$. 13.33 ng μl^{-1} trypsin solution in 9:1 (v/v) 50 mM NH$_4$CO$_3$/50% CH$_3$CN was added and samples were incubated overnight at 37 °C. Peptides were extracted with 1:2 (v/v) 5% formic acid/50% CH$_3$CN and dried. Peptides were desalted by reconstitution in 3:8 (v/v) 70% formic acid/0.1% TFA, followed by loading onto a custom-made stage tip (2 × 1.06 mm punches of Empore C18 extraction disks [3 M] in a 200 μl pipette tip)[41] activated with 80% CH$_3$CN and 0.1% TFA. Tips were washed twice with 0.1% TFA and peptides eluted with 80% CH$_3$CN and 0.1% TFA. Peptides were dried and reconstituted for LC/MS/MS analysis. Capillary LC/MS/MS was carried out using an LTQ Orbitrap Velos (Thermo Scientific) with a nanoAcquity uHPLC (Waters) system as described previously[42]. The data were processed as described previously[11]. MASCOT scores were above

the identity or extensive homology threshold and representative spectra are illustrated in Fig. 2c.

Quantitative assessment of amino acid tolerance. The following workflow was performed to assess the tolerance for natural amino acid substitution at residues chosen for sAA incorporation. Strains were grown to mid-log phase in 1 ml of permissive LB media, and MAGE was performed as described[12] with modifications described here. Post induction of λ-Red proteins, cells were transferred to individual wells of a 96-well, V-bottomed plate, and washed twice at 4 °C with sterile deionized water (dH$_2$O). Cells were re-suspended in 50 μl of water or 1 μM mutagenic single-stranded DNA to convert a single in-frame essential TAG codon to one of 20 sense codons, and electroporated in a 96-well plate. Cells were electroporated using the Harvard BTX electroporation system (2.4 kV, 750 Ω, 25 μF). Electroporated cells were recovered in 1.5 ml of fresh permissive media in a 96-well plate for 4 h at 34 °C shaking. Cells were pelleted, washed twice with sterile dH$_2$O, and re-suspended in 200 μl of 1× PBS. Serial dilutions were made in 1× PBS and 50 μl each of non-diluted and 100-fold diluted samples were plated on solid non-permissive LB media. 50 μl each of higher dilutions were plated on permissive solid media and all plates were incubated for 20 h at 34 °C.

Colony-forming units counted on non-permissive media were expressed as a ratio of total c.f.u. on permissive media. Since the frequency of MAGE-mediated recombination (~0.3)[12] exceeds the escape frequencies of these background strains ($\leq 10^{-5}$), we directly correlated these ratios to amino acid tolerance. MATLAB was used to calculate the $\log_{10}$ of this ratio. Where no c.f.u. were observed on non-permissive media, indicative of a highly intolerant substitution, a ratio could not be calculated and these values were defaulted to NaN within MATLAB. A heat map was used to compare representative data for one experiment, where blue indicates a tolerated substitution and yellow, a non-tolerated substitution.

Twenty-one separate MAGE experiments were performed as described above for each strain, per essential genomic TAG, to assess tolerance for each of the 20 natural amino acids at each TAG site, plus a negative control (water). Strains with one TAG codon (SecY.Y122β, DnaX.Y113β, LspA.Y54β, DnaA.W6γ, SerS.F213γ, and MurG.F243γ) were assessed across 21 (including the negative control) experiments per strain, strains with two TAG codons (rEc.β.dB.9) were assessed across 42 (including two negative control) experiments per strain, and strains with three TAG codons (rEc.β.dC.12 and rEc.γ.dC.46) were assessed across 63 (including three negative control) experiments per strain. Reported results repeated at least three times in independent experiments.

Escape assays. Strains were grown in triplicate to late-log phase in 2 ml of permissive LB media, pelleted, washed twice with sterile dH$_2$O, and re-suspended in 200 μl 1× PBS. To obtain total and escape mutant c.f.u., serial dilutions were made and equal volumes were plated on permissive and non-permissive solid media plates (100 × 15 mm). Plates were incubated at 34 °C and escape frequency was calculated as the total number of escape mutant c.f.u. observed per total cells plated. Reported escape frequencies are means of three technical replicates where error bars represent ±s.d. To isolate escape mutants in strains with lower escape frequencies, ~10^{10}–10^{11} cells were plated and the resulting escape frequency from a representative escape assay is reported. When escape mutants were not detected upon plating ~10^{10}–10^{11} cells, the escape frequency is described to be below the limit of detection and reported as less than a frequency of one over the total number of cells plated. In all cases, reported results repeated at least three times in independent experiments. Where temporal monitoring of escape frequencies on solid media is reported (Fig. 3d), representative escape assays are plotted and results repeated at least three times in independent experiments.

Liquid escape assays. Long-term liquid growth was assessed for two strain backgrounds: the pAzF-dependent strain, rEc.γ.dC.46′.ΔtY, and its non-contained ancestor, rEc.γ. Growth of rEc.γ.dC.46′.ΔtY was separately assessed in permissive (+sAA/+L-arabinose) and non-permissive (−sAA/−L-arabinose) media and growth of rEc.γ was assessed in non-permissive media. In all cases, flasks contained carbenicillin to prevent contamination.

Strains were grown in 100 ml of LB media overnight. Cultures were then pelleted and washed twice with the same volume of sterile dH$_2$O. Washed pellets were re-suspended in LB media plus or minus small molecules and this slurry was then added to shake flasks containing 1 l of LB media plus or minus small molecules, at time zero. At this first time point, a 1 ml sample was obtained from each flask, from which the OD$_{600}$ was measured and 50 μl was plated on both permissive (+sAA/+L-arabinose/+carbenicillin) and non-permissive (−sAA/−L-arabinose/+carbenicillin) solid LB media, in three technical replicates. Average c.f.u. counts are reported and error bars represent ±s.d. In all cases, c.f.u. on solid media were counted after 24 h of incubation at 34 °C. OD$_{600 \text{ nm}}$ readings and c.f.u. counts were collected in this manner for all subsequent time points for the following 20 days.

After 20 days of growth in liquid media, the two 1-l cultures of rEc.γ.dC.46′.ΔtY grown in non-permissive and permissive media were interrogated for the presence of a single escape mutant. The entire culture was pelleted, re-suspended in 7 ml of

1× PBS, and plated across 30 large non-permissive solid media plates that were subsequently monitored for c.f.u. formation over the following 7-day period. All reported results repeated at least three times in independent experiments.

Environmental challenges. Wild-type *E. coli K-12 substr. MG1655*, and additional *E. coli* strains EcNR2[12], rEc.γ.dC.46′ (a pAzF auxotroph), and rEc.γ (non-contained ancestor to rEc.γ.dC.46′) were grown to mid-log phase in 1 ml of LB media supplemented with small molecules, where necessary. Cultures were washed three times with sterile dH$_2$O and re-suspended in 1 ml 1× PBS. A total of 16, two-fold serial dilutions were made and spotted on the following solid media types: (1) LB, (2) EZ rich defined[43] (with modifications by Teknova) and containing 100× carbon source (40% glycerol), (3) blood agar (Teknova), or (4) soil extract agar (HiMedia). Prior to spotting, plates were topically supplemented with pAzF, L-arabinose, and/ or biotin, and dried for at least 1 h. Spotted plates were incubated for 1 day at 34 °C and photographed in a Gel Doc XR+ running ImageLab v4.0.1 (BioRad).

Selectable markers used in this study. *cat* (1,015 bp)

cctgtgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttg
gcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtattt
tttgagttgtcgagatttcaggagctaaggaagctaaaatcactggatataccaccgttgatatatc
ccaatggcatcgtaaagaacatttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggata
ttacggcctttttaaagaccgtaaagaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatga
atgctcatccggaattacgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcaccttgttacaccgtt
ttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcg
caagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgttttcgtctcagccaat
ccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgtttttcaccatgggcaaa
tattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcgg
cagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttaaggcagttattggtgc
cctaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacc
cggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtt

kanR (1,165 bp)

cctgtgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttg
gcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtattt
tttgagttgtcgagatttcaggagctaaggaagctaaaatgagccatattcaacgggaaacgtcgaggccgcgatta
aattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcg
cttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgag
atggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatgg
ttactcaccactgcgatccccgggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttga
tgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtcctttaacagcgatcgcgtatttcgtctcg
ctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaa
caagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataac
cttattttttgacgagggaaattaatagttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgc
catcctatggaactgcctcggtgagtttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgata
tgaataaattgcagtttcatttgatgctcgatgagtttttctaattttttttaaggcagttattggtgcccttaaacgcctggtt
gctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttcagg
gcagggtcgttaaatagccgcttatgtctattgctggtt

spec^R (1,201 bp)

cagccaggacagaaatgcctcgacttcgctgctgcccaaggttgccgggtgacgcacaccgtggaaacggatgaagg
cacgaacccagtggacataagcctgttcggttcgtaagctgtaatgcaagtagcgtatgcgctcacgcaactggtcca
gaaccttgaccgaacgcagcggtggtaacggcgcagtggcggtttttcatggcttgttatgactgtttttttttggggtacag
tctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgtta
cgcagcagggcagtcgccctaaaacaaagttaaacatcatgagggaagcggtgatcgccgaagtatcgactcaacta
tcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtacattacgctccgcagttgcatg
gcggcctgaagccacacagtgatattgatttgctggttacggtgaccgtaaggcttgatgaaacaacgcggcgagcttt
gatcaacgaccttttggaaacttcggctttccctggagagagcgagattctccgcgctgtagaagtcaccattgttgtg
cacgacgacatcattccgtggcgttatccagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattcttg
caggtatcttcgagccagccacgatcgacattgatctggctatcttgctgacaaaagcaagagaacatagcgttgcctt
ggtaggtccagcggcggaggaactctttgatccggttcctgaacaggatctatttgaggcgctaaatgaaacctaacg
ctatggaactcgccgcccgactgggctggcgatgagcgaaatgtagtgcttacgttgtcccgcatttggtacagcgca
gtaaccggcaaaatcgcgccgaaggatgtgctgccgactgggcaatggagcgcctgccggcccagtatcagcccg
tcatacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaat
ttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataaagctttactgagctaataacaggactgctgg
taatcgaggccctttttatttctgca

tolC (1,746 bp)

ttgaggcacattaacgcccctatggcacgtaacgccaccttttgcggtagcggcttctgctagaatccgcaataattttac
agtttgatcgcgctaaatactgcttcaccacaaggaatgcaaatgaagaaattgctccccattcttatcggcctgagcc
tttctgggttcagttcgttgagccaggccgagaacctgatgcaagtttatcgtcaagcacgcctttagtaacccggaatt
gcgtaagtctgccgccatcgtgatgctgcctttgaaaaaatttaatgaagcgcgcagtccattactgccacagctagg
tttaggtgcagattacacctatagcaacggctaccgcgacgcgaacggcatcaactctaacgcgaccagtgcgtcctt
gcagttaactcaatccattttttgatatgtcgaaatggcgtgcgttaacgctgcaggaaaaagcagcagggattcaggac
gtcacgtatcagaccgatcagcaaaaccttgatcctcaacaccgcgaccgcttatttcaacgtgttgaatgctattgacg
ttctttcctatacacaggcacaaaagaagcgatctaccgtcaattagatcaaacatcccaacgttttaacgtggggcct
ggtagcgatcaccgacgttgcaggaaccgcccacagtacgataccgtctggcgaacgaagtgaccgcacgtaat
aaccttgataacgcggtagagcagctgcgccagatcaccgtaactactatccggaactggctgcgctgaatgtcgaa
aactttaaaaccgacaaaccacagccggttaacgcgctgctgaaagaagccgaaaaacgcaacctgtcgctgttaca

ggcacgcttgagccaggacctggcgcgcgagcaaattcgccaggcgcaggatggtcacttaccgactctggatttaa
cggcttctaccgggatttctgacacctcttatagcggttcgaaaacccgtggtgccgctggtacccagtatgacgatag
caatatgggccagaacaaagttggcctgagcttctcgctgccgatttatcagggcggaatggttaactcgcaggtgaa
acaggcacagtacaactttgtcggtgccagcgagcaactggaaagtgcccatcgtagcgtcgtgcagaccgtgcgttc
ctccttcaacaacattaatgcatctatcagtagcattaacgcctacaaacaagccgtagtttccgctcaaagctcattag
acgcgatggaagcgggctactcggtcggtacgcgtaccattgttgatgtgttggatgcgaccaccacgttgtacaacg
ccaagcaagagctggcgaatgcgcgttataactacctgattaatcagctgaatattaagtcagctctgggtacgttgaa
cgagcaggatctgctggcactgaacaatgcgctgagcaaaccggtttccactaatccggaaaacgttgcaccgcaaa
cgccggaacagaatgctattgctgatggttatgcgcctgatagcccggcaccagtcgttcagcaaacatccgcacgca
ctaccaccagtaacggtcataacccttttccgtaactgatgacgacgacggggaagcttaattagctgatctagaggca
tcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtag
gacaaatccgccgccctaga

zeo[R] (762 bp)

Ggtgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggccaag
ttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggctcgggttc
tcccgggacttcgtggaggacgacttcgccggtgtggtccgggacgacgtgaccctgttcatcagcggcggtccaggac
caggtggtgccggacaacaccctggcctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggagg
tcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtggggggcgggagtt
cgccctgcgcgacccggccggccaactgcgtgcacttcgtggccgaggagcaggactgacacgtccgacggcgggcc
acgggtcccaggcctcggagatccgtcccccttttcctttgtcgatatcatgtaattagttatgtcacgcttacattcacg
ccctccccccacatccgctctaaccgaaaaggaaggagttagacaacctgaagtctaggtccctatttattttttttatag
ttatgttagtattaagaacgttatttatatttcaaatttttctttttttctgtacagacgcgtgtacgcatgtaacattatact
gaaaaccttgcttgagaaggttttgggacgctcgaaggctttaatttgcaagct

31. Normally, J., Masson, J. M., Kleina, L. G., Abelson, J. & Miller, J. H. Construction of two *Escherichia coli* amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proc. Natl Acad. Sci. USA* **83,** 6548–6552 (1986).

32. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).

33. Fan, C., Ho, J. M., Chirathivat, N., Soll, D. & Wang, Y. S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* **15,** 1805–1809 (2014).

34. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* **40,** e132 (2012).

35. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli. J. Mol. Biol.* **395,** 361–374 (2010).

36. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).

37. Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. & Blattner, F. R. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27,** 3821–3835 (1999).

38. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4,** 206–223 (2009).

39. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).

40. Rinehart, J. *et al.* WNK2 kinase is a novel regulator of essential neuronal cation-chloride cotransporters. *J. Biol. Chem.* **286,** 30171–30180 (2011).

41. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670 (2003).

42. Heinemann, I. U. *et al.* Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Lett.* **586,** 3716–3722 (2012).

43. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119,** 736–747 (1974).
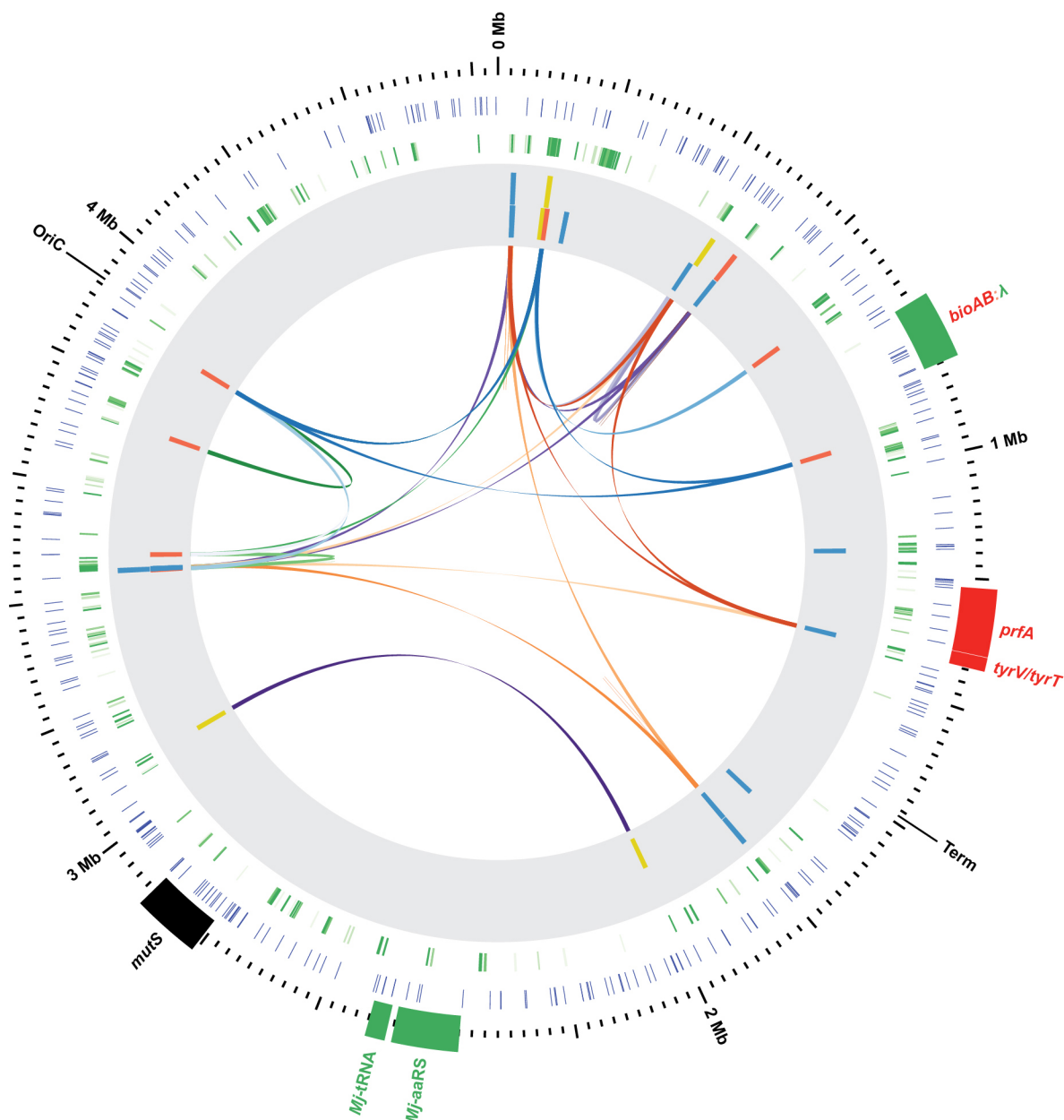
Extended Data Figure 1 | Comprehensive map of synthetic auxotrophs.
Circos plot summarizing synthetic auxotrophs generated in this study. Red and green genes reflect knockouts and insertions, respectively. Outermost ticks indicate genomic location, inner blue ticks indicate locations where TAG codons were converted to TAA in the GRO, and green ticks reflect the locations of 303 *E. coli* essential genes. The shaded grey inner circle contains essential TAG loci in synthetic auxotrophs, where yellow ticks represent amino-terminal insertions, blue ticks represent tolerant substitutions, and red ticks represent functional-site substitutions. Innermost links represent unique combinations of TAGs in higher-order synthetic auxotrophs. Links of a single colour correspond to a single strain.
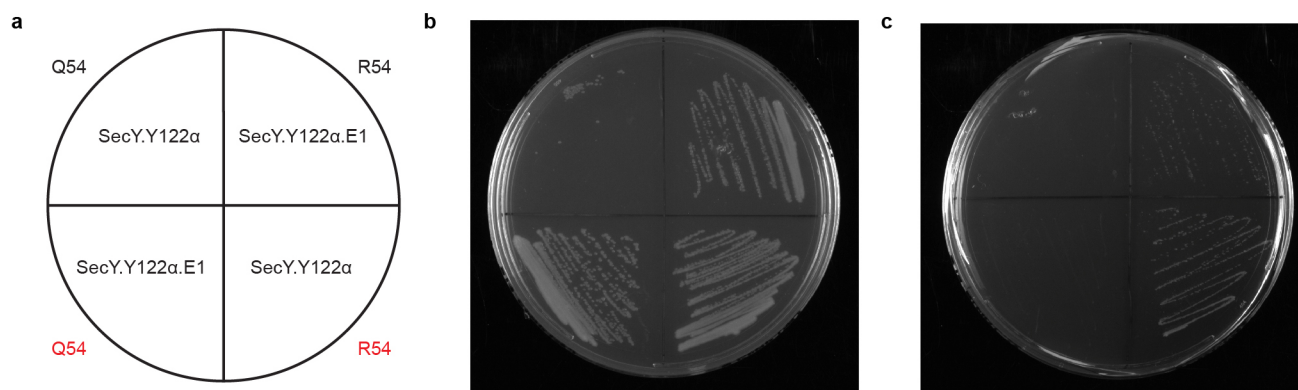
Extended Data Figure 2 | rpsD.Q54R is sufficient for loss of pAcF-dependence in SecY.Y122α. a, Plate map with genotypes of strains shown in **b** and **c**. On the top half of the plate SecY.Y122α.E1 (top right quadrant) contains the *rpsD.Q54R* mutation and is an escape mutant of pAcF-auxotroph, SecY.Y122α (top left quadrant). On the bottom half of the plate the *rpsD.Q54R* mutation was introduced into SecY.Y122α (bottom right quadrant), resulting in a loss of pAcF-dependence, and reverted to wild type in SecY.Y122α.E1 (bottom left quadrant), restoring pAcF-dependence. The amino acid present at residue 54 within RpsD is indicated at the perimeter of the plate, where red signifies that the given mutation was introduced into the genotype by MAGE to demonstrate the causal mechanism of escape. **b**, Growth on solid permissive media demonstrates growth of all strains. **c**, Growth on solid non-permissive media. Introduction of the *rpsD.Q54R* mutation into the synthetic auxotroph SecY.Y122α results in loss of containment (bottom right quadrant). Reverting the mutation to wild type in SecY.Y122α.E1 results in restoration of containment (bottom left quadrant). Together, these data demonstrate that the *rpsD.Q54R* mutation is sufficient for loss of pAcF-dependence in SecY.Y122α.
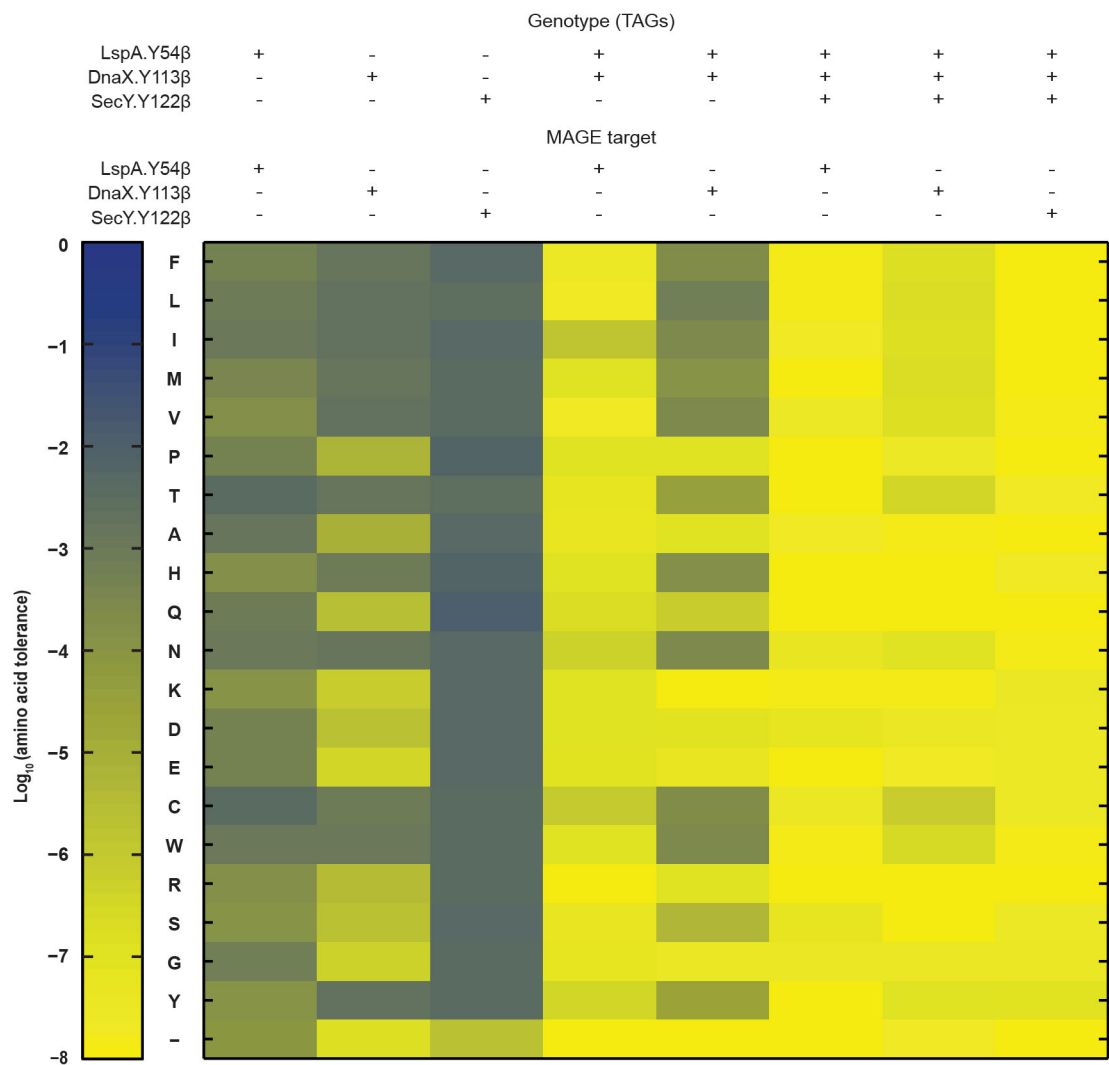
Extended Data Figure 3 | Quantitative assessment of amino acid tolerance in higher-order pIF auxotrophs. Representative assay surveying tolerance of one of three essential TAG loci to the twenty amino acids in different synthetic auxotrophs and expressed as $\log_{10}$ of total cell survival. The + symbol indicates the presence of a TAG codon at the specified locus in the background strain and − indicates the wild-type codon. Blue and yellow indicate high and low tolerance to substitution, respectively. Substitutions DnaX.Y113W and SecY.Y122Q are tolerated but yielded a lower percentage of survival on non-permissive media in a background with two TAGs, an effect that was pronounced in a background with three TAGs. While DnaX.Y113, SecY.Y122 and LspA.Y54 are permissive for most natural amino acids, strains with more than one of these essential TAGs are less prone to survive in the event that any one TAG is compromised. SecY.Y122Q and DnaX.Y113W were tolerated substitutions also observed in real escape mutants of these strains (Supplementary Table 7). Reported results repeated at least three times in independent experiments. Refer to the Methods for a complete description of this experiment.
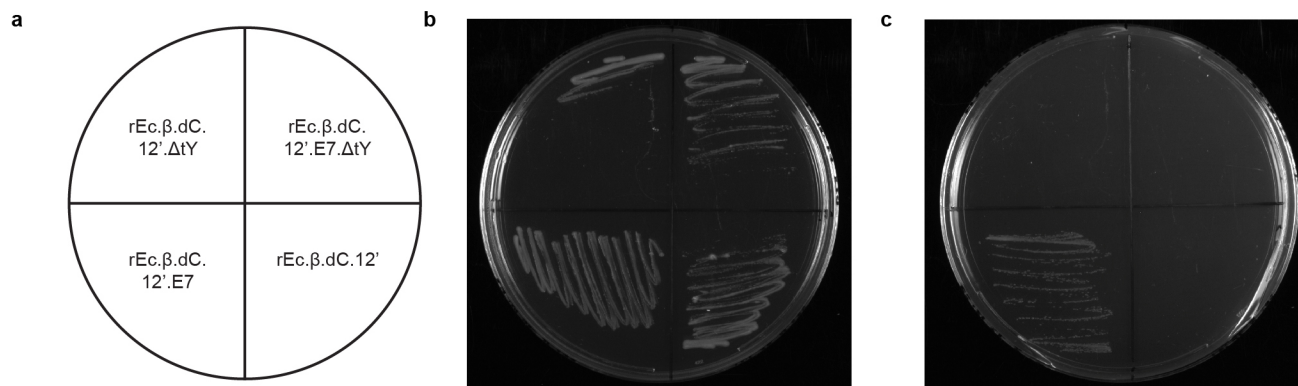
Extended Data Figure 4 | Deletion of *tyrT* and *tyrV* restores pIF-dependence and fitness of rEc.β.dC.12′.E7. **a**, Plate map with genotypes of strains shown in **b** and **c**. rEc.β.dC.12′.E7 is an escape mutant of its sAA-dependent ancestor (rEc.β.dC.12′) and contains a *tyrT* ochre suppressor mutation (*supC*). The fitness of rEc.β.dC.12′.E7 in permissive media is impaired relative to rEc.β.dC.12′, with doubling times of 91.74 (±1.49) and 61.81 (±0.65) minutes, respectively. Tyrosine tRNA redundancy was eliminated (ΔtY) in both strains by λ-Red mediated replacement of *tyrT* and *tyrV* with chloramphenicol acetyltransferase (*cat*), rendering the resulting strains (rEc.β.dC.12′.ΔtY and rEc.β.dC.12′.E7.ΔtY) dependent on *tyrU* for tyrosine incorporation during normal protein synthesis. Elimination of tyrosine redundancy reduced the escape frequency of rEc.β.dC.12′ from 2.17×10^{-9} (Fig. 2e) to $<4.85 \times 10^{-12}$ (no escape mutants were observed upon plating 2.06×10^{11} cells) and restored pIF-dependence in rEc.β.dC.12′.E7 to $<4.73 \times 10^{-12}$ (no escape mutants were observed upon plating 2.12×10^{11} cells). Escape mutants were not detected for either strain up to 7 days after plating on non-permissive media (Fig. 3d and Supplementary Table 11). Tyrosine tRNA deletion also restored the fitness of the escape mutant to approximately that of its sAA-dependent ancestor (60.66 ± 0.12 min). Taken together, these results establish *tyrT* as the causal mechanism of escape in rEc.β.dC.12′.E7. **b**, Growth on solid permissive LB media. **c**, Growth on solid non-permissive LB media. All reported doubling times are averages, where $n = 3$ technical replicates, and error bars represent ±s.d. Refer to the Methods for a complete description of escape frequencies.

Suppressor: None   Trp   Phe   +sAA/+ara: ──────   -sAA/-ara: ·······

a   rEc.γ.dC.46'.ΔtY

b   rEc.γ.dC.46'.ΔtY

Suppressor	Condition	DT (mins)
None	+sAA/+ara	64.91
supU (Trp)	-sAA/-ara	No growth
supPhe (Phe)	-sAA/-ara	252.86

c   rEc.β.dC.12'.ΔtY

d   rEc.β.dC.12'.ΔtY

Suppressor	Condition	DT (mins)
None	+sAA/+ara	60.99
supU (Trp)	-sAA/-ara	380.82
supPhe (Phe)	-sAA/-ara	No growth
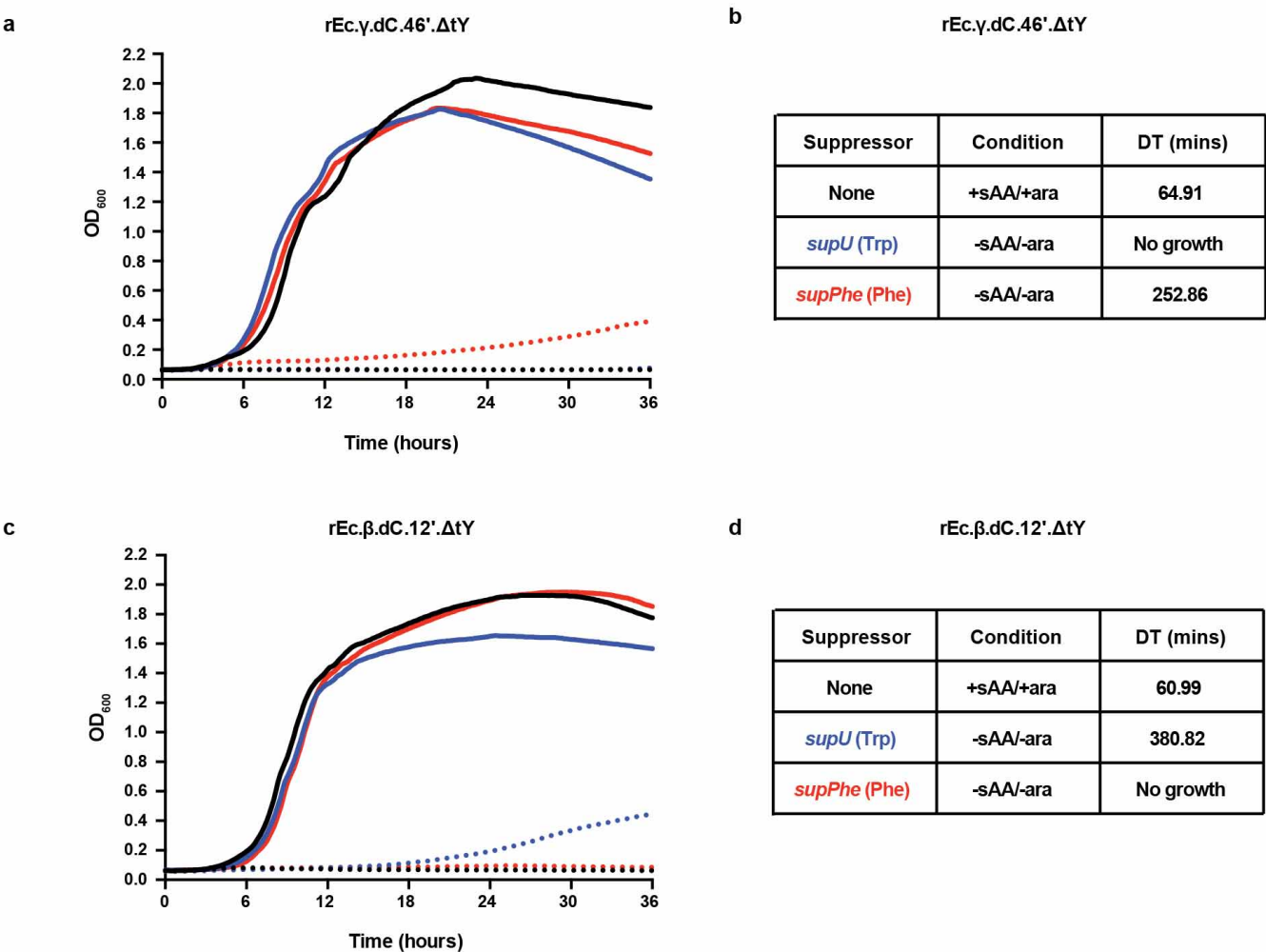
Extended Data Figure 5 | Growth profiles of strains expressing phenylalanine or tryptophan amber-suppressor tRNAs. **a–d**, Growth was assessed for rEc.γ.dC.46′.ΔtY and rEc.β.dC.12′.ΔtY in the presence of amber suppression by either pTech-*supU* (blue), pTech-*supPhe* (red), or in the absence of plasmid-based amber suppression (black). Cells were washed twice with dH$_2$O and re-suspended in the same volume of 1× PBS. Washed cells were normalized by OD$_{600}$ to inoculate roughly equal numbers of cells per well. Growth profiles are shown for rEc.γ.dC.46′.ΔtY (**a**, **b**) and rEc.β.dC.12′.ΔtY (**c**, **d**) in permissive (+sAA/+L-arabinose, solid lines) and non-permissive (−sAA/−L-arabinose, dashed lines) LB liquid media. Doubling times are shown for the ancestral strain (black) in permissive media and suppressor-containing strains (red and blue) in non-permissive media where growth was

observed. Plasmid containing strains were always grown in the presence of zeocin for plasmid maintenance. Growth was never observed for the contained ancestors in non-permissive media (black, dashed lines). In the presence of tryptophan suppression, growth of rEc.γ.dC.46′.ΔtY was not observed and growth of rEc.β.dC.12′.ΔtY was severely impaired (380 min doubling time), with a 6.24-fold increase in doubling time relative to the contained ancestor grown in permissive media. In the presence of phenylalanine suppression, growth of rEc.β.dC.12′.ΔtY was not observed and growth of rEc.γ.dC.46′.ΔtY was severely impaired (252 min doubling time), with a 3.90-fold increase in doubling time relative to the contained ancestor grown in permissive media. Representative growth profiles and doubling times are reported. These results repeated at least three times in individual experiments.
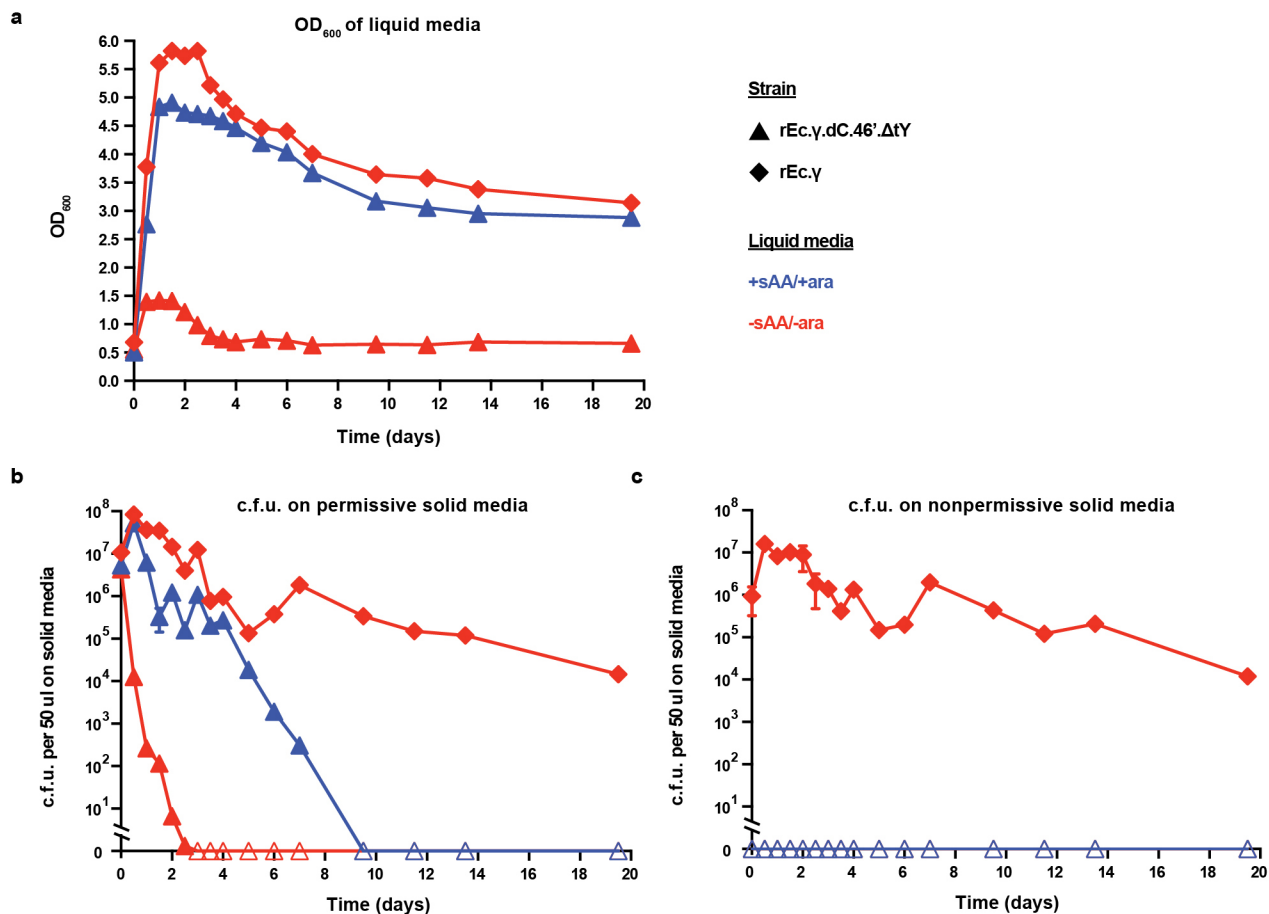
Extended Data Figure 6 | Long-term growth of rEc.γ.dC.46′.ΔtY in liquid LB media relative to rEc.γ. a–c, Approximately 10^{11} cells of strain rEc.γ.dC.46′.ΔtY (indicated with triangles) were inoculated into 1 l of permissive (+sAA/+L-arabinose, blue) or non-permissive (−sAA/−L-arabinose, red) LB media and incubated with agitation at 34 °C for 20 days. Results from the equivalent experiment with the non-contained ancestor rEc.γ (indicated with diamonds) are also shown. Cultures were frequently monitored by OD_{600} (**a**) and quantification of c.f.u. on solid permissive (+sAA/+L-arabinose) (**b**) and non-permissive (−sAA/−L-arabinose) (**c**) LB media. C.f.u. are plotted as the average of three replicates. Open symbols indicate that no c.f.u. were observed. Symbols for rEc.γ.dC.46′.ΔtY are not visible because c.f.u. were never observed from either permissive or non-permissive liquid cultures

plated on non-permissive solid media. At the end of the 20-day growth period, both cultures containing rEc.γ.dC.46′.ΔtY were interrogated for the presence of a single escape mutant by plating each 1 l of culture across 30 non-permissive solid media plates. C.f.u. were not observed and remained below the limit of detection for the following 7-day observation period. We hypothesize that the decrease in c.f.u. counts obtained on permissive solid media for the permissive culture of rEc.γ.dC.46′.ΔtY reflects pAzF degradation at ≥6 days. Reported c.f.u. values are averages, where $n = 3$ technical replicates, and error bars are ±s.d. Reported results repeated at least three times in independent experiments. Refer to the Methods for a complete description of this experiment.
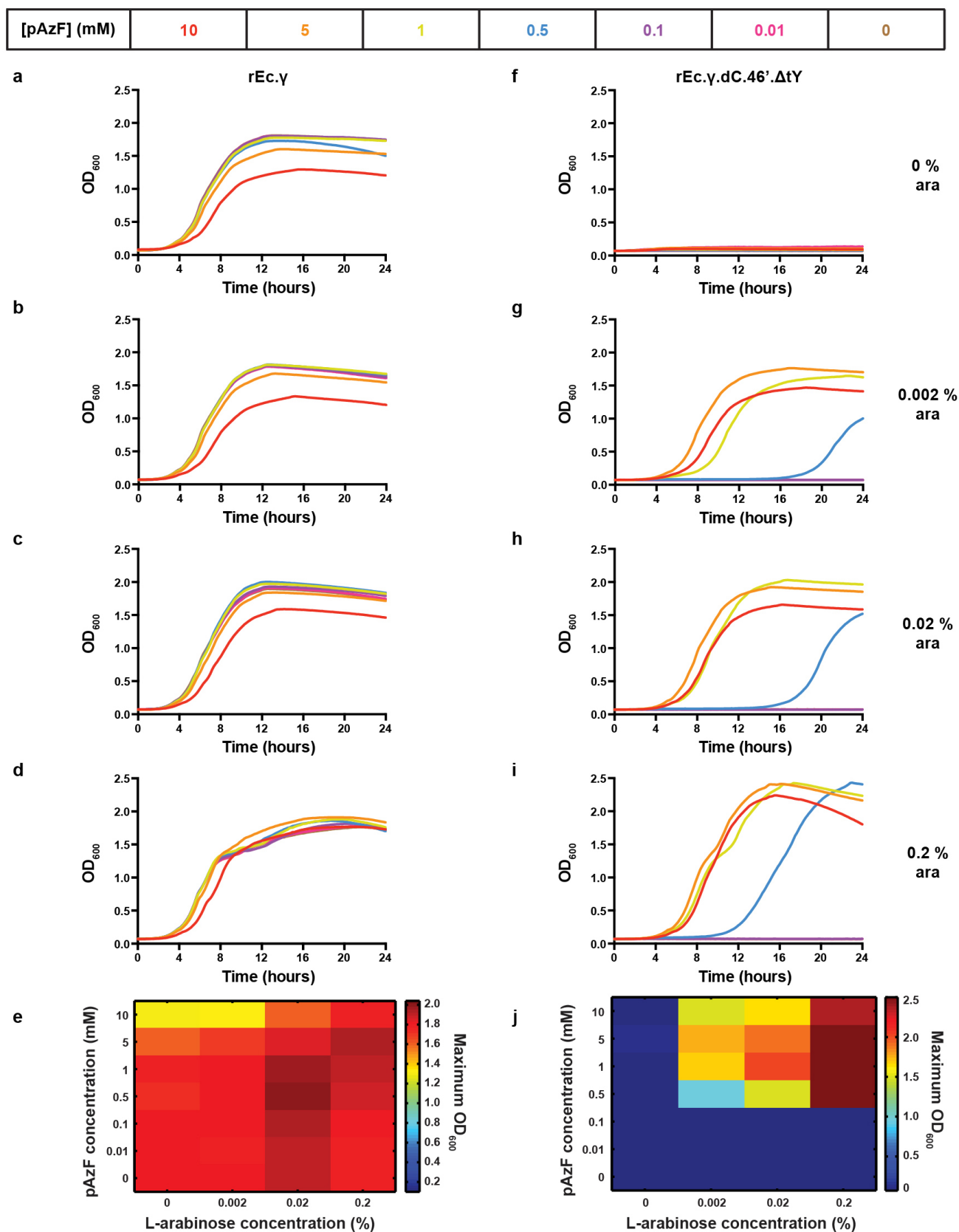
Extended Data Figure 7 | Dose-dependent growth of rEc.γ.dC.46′.ΔtY in pAzF and L-arabinose compared to the non-contained ancestor. Growth in LB media supplemented with different concentrations of pAzF and L-arabinose. **a–d**, Growth profiles for rEc.γ across a gradient of pAzF concentrations in the presence of 0% (**a**), 0.002% (**b**), 0.02% (**c**) and 0.2% (**d**) L-arabinose. **f–i**, Growth profiles for rEc.γ.dC.46′.ΔtY across a gradient of pAzF

concentrations in the presence of 0% (**f**), 0.002% (**g**), 0.02% (**h**) and 0.2% (**i**) L-arabinose. **e, j**, Growth profiles illustrated in **a–d** and **f–i** are depicted as heat maps in **e** and **j**, respectively, where the maximum OD$_{600}$ was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where $n = 3$ technical replicates. Reported results repeated at least three times in independent experiments.

[pIF] (mM)	10	5	1	0.5	0.1	0.01	0
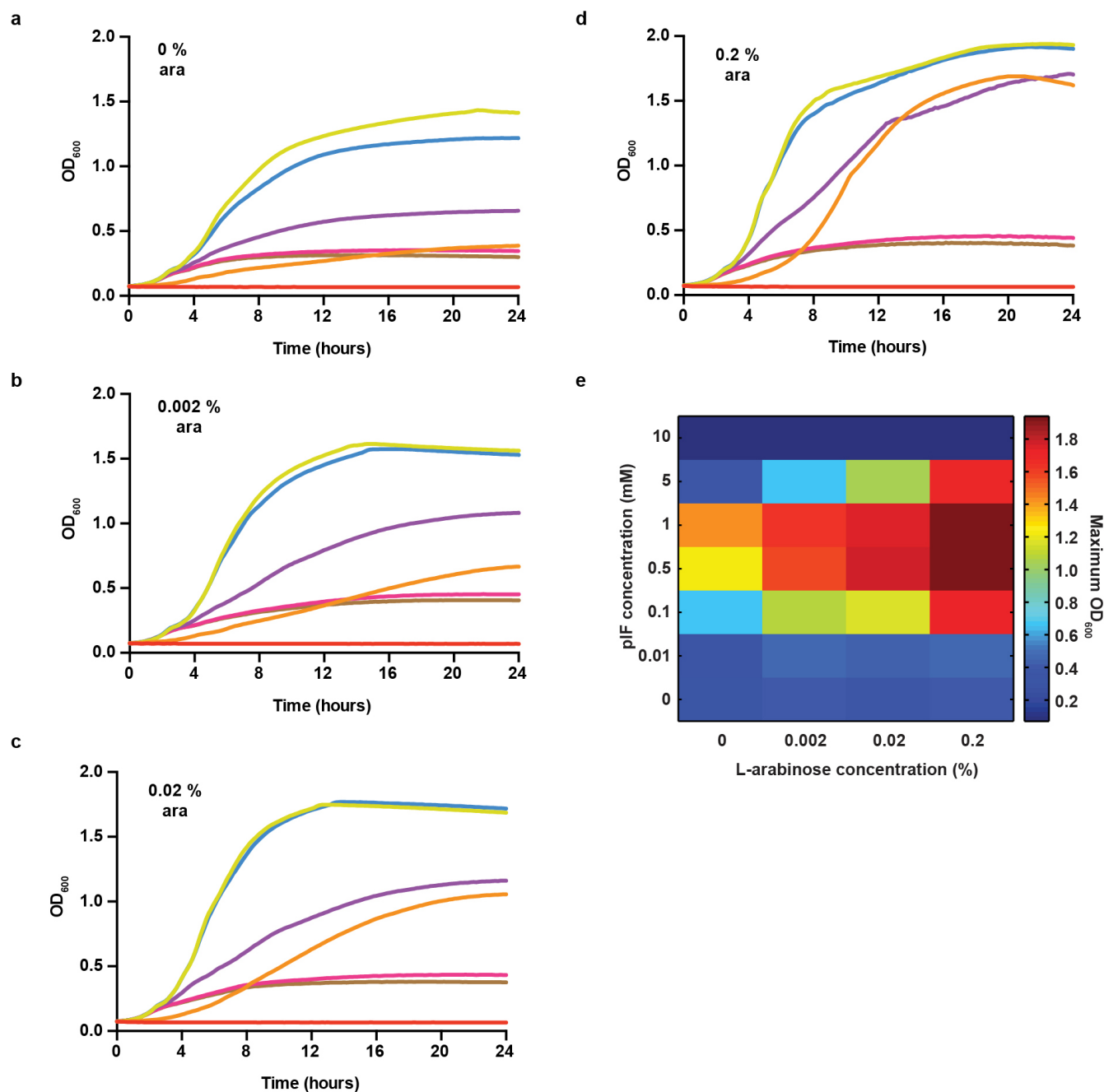
Extended Data Figure 8 | Dose-dependent growth of rEc.β.dC.12′.ΔtY in pIF and L-arabinose. Growth in LB media supplemented with different concentrations of pIF and L-arabinose. **a–d**, Growth profiles for rEc.β.dC.12′.ΔtY across a gradient of pIF concentrations in the presence of 0% (**a**), 0.002% (**b**), 0.02% (**c**) and 0.2% (**d**) L-arabinose. **e**, Growth profiles illustrated in parts **a–d** are depicted as a heat map, where the maximum OD$_{600}$ was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where $n = 3$ technical replicates. Reported results repeated at least three times in independent experiments.
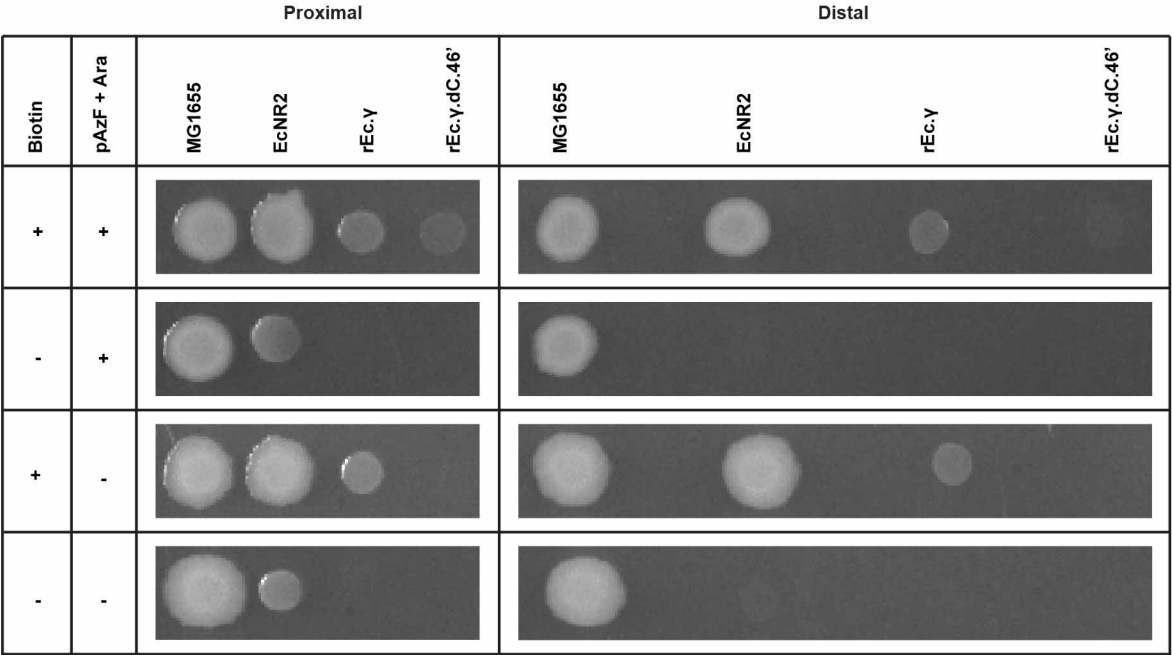
Extended Data Figure 9 | Proximity-dependent complementation of biotin auxotrophy. Wild-type *E. coli K-12 substr. MG1655* and three strains auxotrophic for biotin, EcNR2, rEc.γ (a non-contained GRO with an integrated pAzF OTS) and rEc.γ.dC.46′ (also a synthetic auxotroph) were grown either adjacent or separately on rich-defined solid media. EcNR2 grew on biotin-deficient media when plated in close proximity to wild-type *E. coli*, suggesting cross-feeding of the essential metabolite. The pAzF auxotroph only grew on media supplemented with biotin, pAzF and L-arabinose.

Predicting climate–driven regime shifts versus rebound potential in coral reefs

Nicholas A. J. Graham[1], Simon Jennings[2,3], M. Aaron MacNeil[1,4], David Mouillot[1,5] & Shaun K. Wilson[6,7]

Climate-induced coral bleaching is among the greatest current threats to coral reefs, causing widespread loss of live coral cover[1]. Conditions under which reefs bounce back from bleaching events or shift from coral to algal dominance are unknown, making it difficult to predict and plan for differing reef responses under climate change[2]. Here we document and predict long-term reef responses to a major climate-induced coral bleaching event that caused unprecedented region-wide mortality of Indo-Pacific corals. Following loss of >90% live coral cover, 12 of 21 reefs recovered towards pre-disturbance live coral states, while nine reefs underwent regime shifts to fleshy macroalgae. Functional diversity of associated reef fish communities shifted substantially following bleaching, returning towards pre-disturbance structure on recovering reefs, while becoming progressively altered on regime shifting reefs. We identified threshold values for a range of factors that accurately predicted ecosystem response to the bleaching event. Recovery was favoured when reefs were structurally complex and in deeper water, when density of juvenile corals and herbivorous fishes was relatively high and when nutrient loads were low. Whether reefs were inside no-take marine reserves had no bearing on ecosystem trajectory. Although conditions governing regime shift or recovery dynamics were diverse, pre-disturbance quantification of simple factors such as structural complexity and water depth accurately predicted ecosystem trajectories. These findings foreshadow the likely divergent but predictable outcomes for reef ecosystems in response to climate change, thus guiding improved management and adaptation.

Some mass bleaching events have resulted in the loss of almost all live coral within individual nations[3]. Examples of coral recovery following severe bleaching have been documented[4,5], yet theory predicts that regime shifts to new benthic assemblages, such as macroalgae, are also likely[2]. To date, there are no documented coral reef regime shifts attributed specifically to climate change. Evidence for coral reef regime shifts due to other causes comes almost exclusively from the Caribbean[6,7], with limited knowledge from the more extensive and diverse Indo-Pacific reef province[8]. Ongoing uncertainty about Indo-Pacific coral reef responses to climate impacts has generated considerable debate regarding appropriate management and adaptation plans, especially those related to no-take marine reserves[9,10], that can be resolved through understanding site and ecosystem characteristics that dictate reef ecosystem trajectories.

Using a 17-year data set, spanning a major climate-induced bleaching event, we assess the long-term ecosystem dynamics of 21 reef sites across the inner islands of Seychelles. Seychelles reefs were the most severely affected globally by the 1998 coral bleaching event, in which a strong El Niño coincided with the Indian Ocean dipole[11]. Across all sites before the bleaching event, average hard coral cover was 28% and macroalgal cover 1% (Fig. 1a), within average bounds for the Indian Ocean region at the time[12]. The mass bleaching event was severe across all reefs in the inner Seychelles, with coral cover reduced by >90%[13]. Both hard coral cover and macroalgal cover steadily increased between 2005 and 2011,

with high heterogeneity reflecting markedly different trajectories among sites (Fig. 1a, Extended Data Figs 1 and 2).

Using four complimentary metrics (Methods), we defined a recovering reef as having greater post-disturbance coral cover than macroalgae cover, with coral cover remaining high or increasing. A regime shifting reef was defined as having greater post-disturbance macroalgal cover than coral cover, with macroalgal cover remaining high or increasing. We used an index, based on the Euclidian distance from pre-bleaching benthic composition, to visualize the differing benthic trajectories of
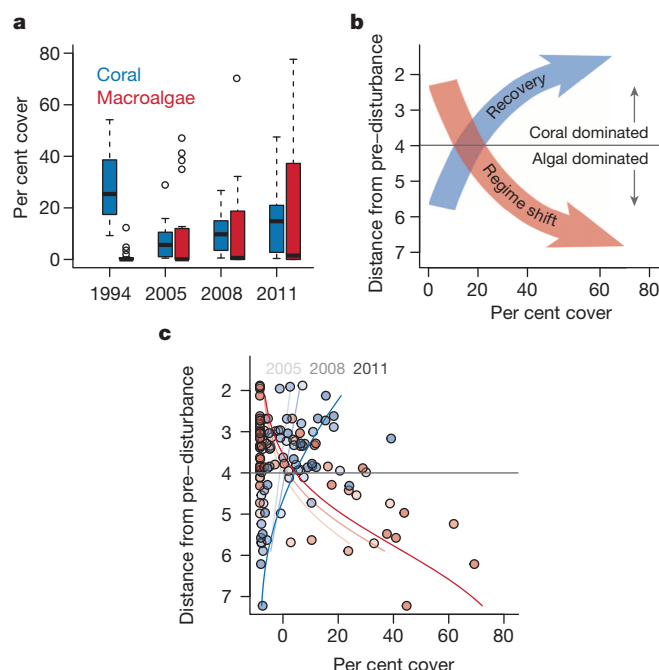

Figure 1 | Recovery and regime shift dynamics on Seychelles coral reefs. **a**, Box (median and 50% quantile) and whisker (95% quantile) plots of coral cover and macroalgal cover in 1994 and through time after the 1998 bleaching event ($n = 84$). **b**, Conceptual diagram for visualizing recovery and regime shift dynamics, using post-disturbance cover of coral (blue) and macroalgae (red) and Euclidian distance of benthic composition from pre-disturbance values. **c**, Recovery and regime shifts on Seychelles reefs. Lines are multinomial model fits for percent coral and macroalgae on each of the 21 reef sites (years distinguished by shading), with Euclidian distance from pre-disturbance composition as a predictor ($n = 63$). Each reef site is represented by two points, one for coral cover (blue dots) and one for macroalgal cover (red dots). Twelve reefs (top of figure) are recovering, moving towards pre-disturbance (1994) compositions, with increasing coral cover and low macroalgal cover, whereas 9 reefs are regime shifting, with increasing distance from 1994 compositions, increasing macroalgal cover and low coral cover (Extended Data Table 1, Extended Data Fig. 4).

[1]Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811 Australia. [2]Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft NR33 OHT, UK. [3]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [4]Australian Institute of Marine Science, PMB 3 Townsville MC, Townsville, Queensland 4810, Australia. [5]ECOSYM, UMR CNRS-UM2 5119, Université Montpellier 2, 34095 Montpellier Cedex, France. [6]Department of Parks and Wildlife, Kensington, Perth, Western Australia 6151, Australia. [7]School of Plant Biology, Oceans Institute, University of Western Australia, Crawley, Western Australia 6009, Australia.

regime shifting and recovering reefs (Fig. 1b). Of the 21 reefs surveyed, 12 followed a post-bleaching recovery trajectory, with live coral cover increasing through time to an average cover of 23% by 2011, macroalgal cover remaining low (<1%), and benthic composition moving towards that observed in 1994 (Fig. 1c, Extended Data Table 1, Extended Data Figs 3 and 4). Annual increases in coral cover on recovering reefs were low for 7–10 years following bleaching, but then increased substantially, probably reflecting increasing local recruitment levels necessary for rapid recovery in isolated coral communities[4]. In contrast, 9 of the 21 reefs followed a post-bleaching regime shift trajectory, with average macroalgal cover steadily increasing to 42% by 2011, coral cover remaining low (<3%), and benthic composition diverging substantially from the pre-bleaching state (Fig. 1c, Extended Data Table 1, Extended Data Figs 3 and 4). Percent cover of coral and macroalgae did not differ between these two groups of reefs in 1994 (Extended Data Table 1), providing the first clear evidence that regime shifts to macroalgae-dominated states occur in response to major coral bleaching events on Indo-Pacific reefs.

Divergent ecosystem trajectories following bleaching are likely to have major implications for reef associated organisms. The functional structure of communities, which captures species' roles based on their biological traits, is predicted to show deterministic links to disturbance[14], but predictions are rarely tested with data. Here we show changes to functional structure of reef fish assemblages are strongly tied to the post-bleaching benthic response (Fig. 2, Methods). Seven years after bleaching, assemblages on all reefs had fewer small bodied species and a less complex functional structure, typified by increasing dominance of invertebrate feeding fishes. On recovering reefs, a smaller initial change in functional structure was followed by a return towards the pre-bleaching state (Fig. 2b), whereas the much greater change on reefs that underwent regime shifts progressed in post-bleaching years, suggesting important functions are being eroded (Fig. 2c).

We assessed the association between 11 reef-level factors and ecosystem trajectories post-disturbance, a priori selected for their generic importance in structuring coral reef ecosystems globally (Extended Data Table 2). Five of the factors correctly (97% of the time) characterized post-bleaching reef trajectories as 'recovery' or 'regime shift': density of juvenile corals, initial structural complexity of the reef, water depth, biomass of herbivorous fishes (also reflecting herbivore diversity, which was collinear), and nutrient conditions of the reef (Fig. 3). Estimates of critical values for each factor, where recovery was more likely to occur than a regime shift, showed that juvenile coral densities >6.2 per m^2 reduce the probability of a regime shift (Fig. 3a), consistent with the expected role of coral recruitment and survival in reef recovery[4]. Coral larval supply does not differ among recovering and regime shifting reefs in Seychelles, but post-settlement survival is lower on regime shifted reefs[15], likely

due to unsuitable settlement substrate, or enhanced post-settlement mortality due to macroalgal overgrowth or alleopathic mechanisms[2,16]. Structural complexity, which captures the structure provided by corals and the underlying reef matrix, decreased the risk of a regime shift trajectory when values before disturbance were >3.1 (widespread moderately complex relief) (Fig. 3b). Structural complexity influences a range of ecological processes, being a substantial contributor to diversity and productivity of many reef associated organisms[17,18]. Sites deeper than 6.6 m depth were less likely to undergo a regime shift (Fig. 3c), possibly reflecting the relationship between light penetration and algal growth[19], or greater vulnerability of shallower reefs to disturbances such as recurrent coral bleaching or storm damage[20]. Experimental and modelling studies will be necessary to clarify the principal mechanisms by which structural complexity and depth influence ecosystem trajectories.

Reef fish herbivory is a key process mediating competition between corals and algae[21], and herbivore biomass, implicitly linked to body size, is a good proxy for overall rates of herbivory[22]. Interestingly, a relatively low biomass of herbivores (177 kg ha^{-1})—below average values for the Indian Ocean[23]—reduced the risk of a regime shift occurring (Fig. 3d). Algal proliferation and dominance over corals is also influenced by nutrient input to reef systems[24]. Here, lower C:N ratios in macroalgal fronds, indicative of higher nutrient loads[25], were linked to a greater likelihood of regime shifts, with the likelihood of regime shifts reducing below 50% when ratios passed 38 (Fig. 3e). Previous debate has focused on whether levels of herbivory or nutrients mediate coral reef regime shifts or recovery[26,27]. Our results suggest that although both variables relate to ecosystem trajectory, they are weaker and less certain predictors than structural complexity, depth, and the density of juvenile corals.

Reefs within no-take marine reserves were no more likely to recover than reefs in fished areas. Although marine reserves may have a positive long-term influence on coral cover in the absence of disturbance[28], our results suggest they may have little influence on post-disturbance benthic trajectories. In Seychelles, marine reserves do enhance herbivore biomass (2005 values: 279 kg ha^{-1} ± 21.5 s.e.) compared to fished areas (mean 163 kg ha^{-1} ± 58.6 s.e.), but biomass in fished areas was still close to the threshold we identified for recovery. Marine reserves may have a greater role in aiding coral recovery in nations where herbivorous fishes are more heavily exploited or where fishing gears that reduce structural complexity are used.

Collecting data on nutrients and juvenile coral densities can be challenging for the often resource-limited agencies charged with monitoring and managing coral reef ecosystems, especially over large areas. We therefore assessed the predictive power of structural complexity and water depth alone to predict reef trajectories. Both variables can be rapidly recorded over large areas and are relatively stable through time, allowing extensive pre-disturbance data to be compiled. Using only these two
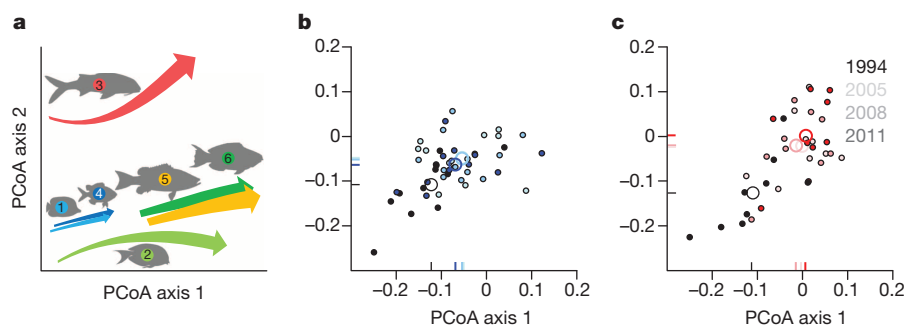

Figure 2 | Trajectories in the functional structure of fish assemblages.
a, Functional trait space underlying the analyses, with arrows depicting increasing body size amongst each fish functional group, depicted by fish outlines: corallivores (1), grazing herbivores (2), invertebrate feeders (3), planktivores (4), piscivores (5), scraping and excavating herbivores (6). **b, c,** Position of recovered (**b**) and regime shifted (**c**) reef sites in the functional space. Black dots represent the mean trait values in 1994, pale to darker colours

represent 2005–2011 ($n = 84$). Large rings indicate centroid for sites within each year (2005 and 2008 overlap in **b**). Dashes on the axes indicate centroid positions for each year. Year and ecosystem trajectory were significant factors (Methods, ANOVA model: axis 1 year $P < 0.001$, trajectory $P < 0.01$; axis 2 year $P < 0.001$, trajectory $P < 0.05$). Fish graphics by T. Saxby, D. Kleine and J. Woerner (Integration and Application Network, University of Maryland Center for Environmental Science, http://ian.umces.edu/imagelibrary/).
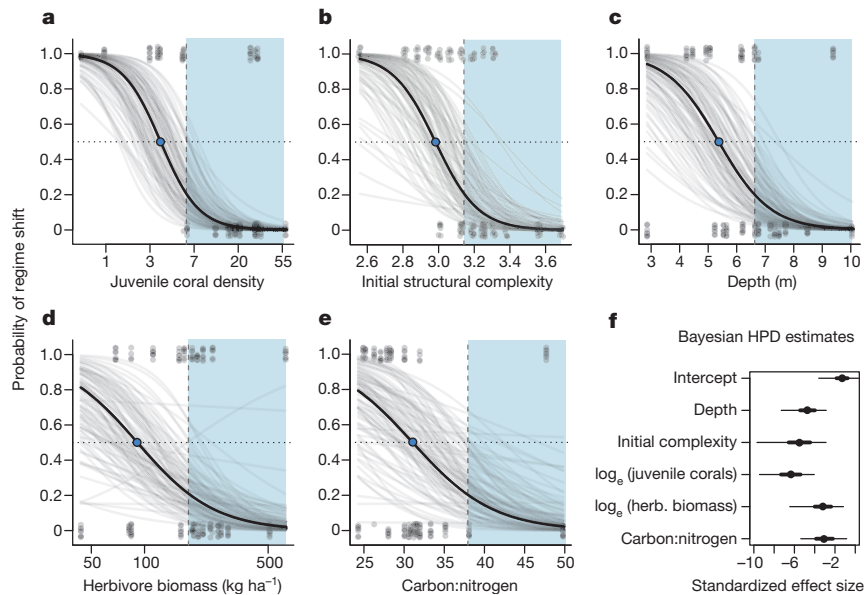
Figure 3 | Bayesian hierarchical logistic regression of predictor variables for the probability of a regime shift. a–e, Marginal plots for predictor variables included in the full model: solid black line represents the mean model fit; grey lines are random resamples of the parameter estimates, representing variability of the expected model fit. Grey dots represent replicates from regime shifted (1) or recovering (0) reefs ($n = 184$). The blue dot represents the point at which regime shifts and recovery are equally likely. Blue shading is beyond upper 95% uncertainty interval for P (shift) = 0.5, where regime shifts are decreasingly likely. **f.** Effect size medians (black dot), with 50% (thick line) and 95% (thin line) highest posterior density (HPD) uncertainty intervals.

variables, we correctly predicted reef trajectory 98% of the time. This suggests that, when combined, structural complexity and depth can effectively identify reefs with a high likelihood of recovery or regime shift, thereby informing reef risk assessment and guiding marine spatial planning initiatives (Fig. 4).

We have demonstrated persistent and divergent responses of coral reefs to climate-induced bleaching with potential consequences for reef fish functioning. With predictions that mass coral bleaching events will increase in frequency[1], our findings foreshadow the likely divergent trajectories expected on other Indo-Pacific reefs. We show that several factors can affect reef ecosystem trajectories following bleaching but, where

necessary, just depth and structural complexity may be useful predictors of ecosystem fate. The predictor factors we identify are important for coral reefs globally (Methods) and the range of values reported for Seychelles is consistent with other locations in the Indo-Pacific (Extended Data Table 3). Further, depth and initial structural complexity evaluated across 6 other countries from East Africa to the South Pacific were consistent predictors of coral or macroalgal dominated reefs post-disturbance, with effect sizes overlapping those from Seychelles (Extended Data Fig. 5). However, there may be regional variations in the relative contribution of factors and threshold values we identified that require more investigation. Uncovering the predictors that dictate reef trajectories following major bleaching events can inform ecosystem management and strategies for human adaptation. Mapping probable ecosystem trajectories can enable limited management resources to target actions, such as phasing out fishing gears that cause habitat damage[29], on reefs where increased structural complexity could help promote recovery[29]. Spatial understanding of ecosystem vulnerability to climatic impacts also holds great potential to be linked with social vulnerability assessments to develop the human adaptation strategies needed to cope with anticipated regime shifts and associated changes in ecosystem services[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.
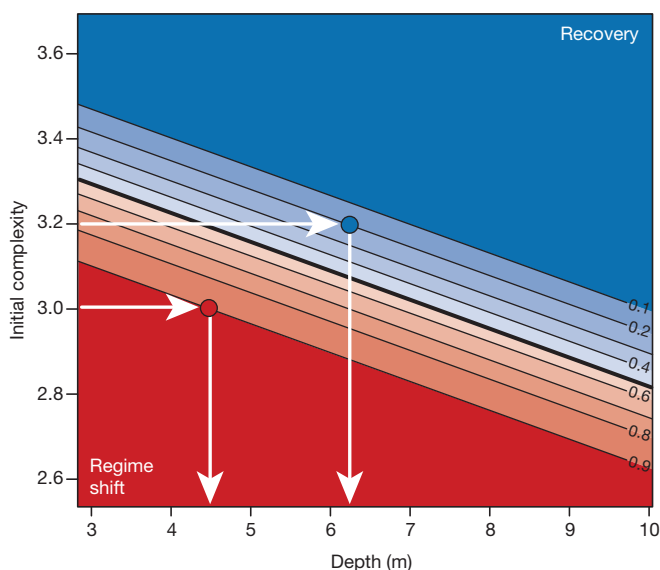
Figure 4 | Contour biplot of the probability of regime shift (red shading) or recovery (blue shading) based on initial structural complexity and water depth. For example, for a complexity value of 3, there is a 0.9 probability of a regime shift occurring for sites <4.5 m depth (red dot), and only sites deeper than ~8 m are more likely to recover. For a complexity value of 3.2, sites deeper than 6.3 m are highly likely to recover (blue dot).

1. Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318,** 1737–1742 (2007).
2. Hughes, T. P., Graham, N. A. J., Jackson, J. B. C., Mumby, P. J. & Steneck, R. S. Rising to the challenge of sustaining coral reef resilience. *Trends Ecol. Evol.* **25,** 633–642 (2010).
3. Goreau, T., McClanahan, T., Hayes, R. & Strong, A. Conservation of coral reefs after the 1998 global bleaching event. *Conserv. Biol.* **14,** 5–15 (2000).
4. Gilmour, J. P., Smith, L. D., Heyward, A. J., Baird, A. H. & Pratchett, M. S. Recovery of an isolated coral reef system following severe disturbance. *Science* **340,** 69–71 (2013).
5. Sheppard, C. R. C., Harris, A. & Sheppard, A. L. S. Archipelago-wide coral recovery patterns since 1998 in the Chagos Archipelago, central Indian Ocean. *Mar. Ecol. Prog. Ser.* **362,** 109–117 (2008).
6. Hughes, T. P. Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. *Science* **265,** 1547–1551 (1994).

7.   Mumby, P. J. *et al.* Fishing, trophic cascades, and the process of grazing on coral reefs. *Science* **311,** 98–101 (2006).
8.   Roff, G. & Mumby, P. J. Global disparity in the resilience of coral reefs. *Trends Ecol. Evol.* **27,** 404–413 (2012).
9.   McCook, L. J. *et al.* Adaptive management of the Great Barrier Reef: A globally significant demonstration of the benefits of networks of marine reserves. *Proc. Natl Acad. Sci. USA* **107,** 18278–18285 (2010).
10. Mora, C. & Sale, P. F. Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Mar. Ecol. Prog. Ser.* **434,** 251–266 (2011).
11. Saji, N. H., Goswami, B. N., Vinayachandran, P. N. & Yamagata, T. A dipole mode in the tropical Indian Ocean. *Nature* **401,** 360–363 (1999).
12. Ateweberhan, M., McClanahan, T. R., Graham, N. A. J. & Sheppard, C. R. C. Episodic heterogeneous decline and recovery of coral cover in the Indian Ocean. *Coral Reefs* **30,** 739–752 (2011).
13. Graham, N. A. J. *et al.* Dynamic fragility of oceanic coral reef ecosystems. *Proc. Natl Acad. Sci. USA* **103,** 8425–8429 (2006).
14. Mouillot, D., Graham, N. A. J., Villéger, S., Mason, N. W. H. & Bellwood, D. R. A functional approach reveals community responses to disturbances. *Trends Ecol. Evol.* **28,** 167–177 (2013).
15. Chong-Seng, K. M., Graham, N. A. J. & Pratchett, M. S. Bottlenecks to coral recovery in the Seychelles. *Coral Reefs* **33,** 449–461 (2014).
16. Smith, J. E. *et al.* Indirect effects of algae on coral: algae-mediated, microbe-induced coral mortality. *Ecol. Lett.* **9,** 835–845 (2006).
17. Graham, N. A. J. & Nash, K. L. The importance of structural complexity in coral reef ecosystems. *Coral Reefs* **32,** 315–326 (2013).
18. Rogers, A., Blanchard, J. L. & Mumby, P. J. Vulnerability of coral reef fisheries to a loss of structural complexity. *Curr. Biol.* **24,** 1000–1005 (2014).
19. McCook, L. J. Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef. *Coral Reefs* **18,** 357–367 (1999).
20. Bridge, T. C. L., Hughes, T. P., Guinotte, J. M. & Bongaerts, P. Call to protect all coral reefs. *Nature Climate Change* **3,** 528–530 (2013).
21. Bellwood, D. R., Hughes, T. P., Folke, C. & Nystrom, M. Confronting the coral reef crisis. *Nature* **429,** 827–833 (2004).
22. Lokrantz, J., Nyström, M., Thyresson, M. & Johansson, C. The non-linear relationship between body size and function in parrotfishes. *Coral Reefs* **27,** 967–974 (2008).
23. McClanahan, T. R. *et al.* Critical thresholds and tangible targets for ecosystem-based management of coral reef fisheries. *Proc. Natl Acad. Sci. USA* **108,** 17230–17233 (2011).
24. Lapointe, B. E. Nutrient thresholds for bottom-up control of macroalgal blooms on coral reefs in Jamaica and southeast Florida. *Limnol. Oceanogr.* **42,** 1119–1131 (1997).
25. Atkinson, M. J. & Smith, S. V. C:N:P ratios of benthic marine plants. *Limnol. Oceanogr.* **26,** 1074–1083 (1983).
26. Smith, J. E., Smith, C. M. & Hunter, C. L. An experimental analysis of the effects of herbivory and nutrient enrichment on benthic community dynamics on a Hawaiian reef. *Coral Reefs* **19,** 332–342 (2001).
27. Burkepile, D. E. *et al.* Chemically mediated competition between microbes and animals: microbes as consumers in food webs. *Ecology* **87,** 2821–2831 (2006).
28. Selig, E. R. & Bruno, J. F. A global analysis of the effectiveness of marine protected areas in preventing coral loss. *PLoS ONE* **5,** e9278 (2010).
29. Cinner, J. E. *et al.* Gear-based fisheries management as a potential adaptive response to climate change and coral mortality. *J. Appl. Ecol.* **46,** 724–732 (2009).
30. Adger, W. N., Hughes, T. P., Folke, C., Carpenter, S. R. & Rockström, J. Social-ecological resilience to coastal disasters. *Science* **309,** 1036–1039 (2005).

Author Contributions N.A.J.G. conceived of the study with S.K.W. and M.A.M.; N.A.J.G. S.J., and S.K.W. collected the data; N.A.J.G., M.A.M., and D.M. developed and implemented the analyses; N.A.J.G. led the manuscript with S.J., M.A.M., D.M., and S.K.W.

METHODS

Field Surveys and indices. Twenty one reefs throughout the inner Seychelles Islands were surveyed using identical methods in 1994, before the mass coral bleaching of 1998, and in 2005, 2008 and 2011, following the bleaching event. The surveys incorporated three different reef habitat types; carbonate fringing reefs; granitic rocky reefs with coral growth; and patch reef habitats on sand, rubble, or rock base. Eight to sixteen replicate 7 m radius point counts were surveyed along the reef slope on each reef, covering up to 0.5 km of reef front and 2,500 m^2 of reef habitat. The underlying substrate type and water depth were recorded at each site. Within each point count area, the percent cover of different growth forms of live hard coral, soft coral, macroalgae, sand, rubble, and rock was quantified, and the structural complexity of the reef was visually estimated on a 6 point scale[13]. This structural complexity measure captures landscape complexity, including the complexity provided by live corals, that of the underlying reef matrix and other geological features, and has been shown to correlate well to other measures of complexity, such as measures of reef height and the linear versus contour chain method[31]. The complexity of each individual count area was assigned to one of the following categories: 0 = no vertical relief, flat or rubbly areas; 1 = low (<30 cm high) and sparse relief; 2 = low but widespread relief; 3 = widespread moderately complex (30–60 cm high) relief; 4 = widespread very complex (60–100 cm high) relief with numerous fissures and caves; 5 = exceptionally complex (>1 m high) relief with numerous caves and overhangs. Estimates can be standardised among observers with minimal training, thereby providing reliable and comparable data of ecological relevance[31]. The density and individual sizes of 134 species of diurnally active, non-cryptic, reef associated fish were recorded within each point count area. Length estimation was calibrated at the beginning of each day by estimating the length of known lengths of PVC pipe. Accuracy was within 4% of actual lengths[32]. We converted data on fish counts to biomass with published length–weight relationships[33,34]. Fish were assigned to feeding groups based on their dominant diets and feeding behaviour[35]. The density of juvenile corals (<10 cm diameter) was quantified in 2011, within 8 replicate 33 cm by 33 cm randomly placed quadrats in each point count area. Juvenile coral data were also collected in 2008, and correlated strongly across sites with 2011 data ($r = 0.85$), suggesting that patterns in juvenile coral densities are fairly consistent through the post-disturbance years surveyed. Sea urchin density (family: Diadematidae) was counted in 2008 in each point count area. Meta-data from these surveys will be deposited in the James Cook University research data repository, the Tropical Data Hub (https://eresearch.jcu.edu.au/tdh).

To assess local nutrient levels ten fronds from each of ten different *Sargassum* thalli were taken from each reef site in 2014. These samples were dried for 48 h at 60 °C, powdered and analysed for nitrogen and carbon abundance. Carbon and nitrogen weight percent (%C and %N, respectively), and C:N ratios were determined using a Costech Elemental Analyzer fitted with a zero-blank auto-sampler at James Cook University's Advanced Analytical Centre, Cairns. The ratio of carbon to nitrogen in macroalgae is a stable, longer-term indicator of ambient nutrient regimes at a given location and should reflect stable spatial patterns among sites[25,36]. Specifically, nutrient inputs incorporated into algal tissue are averaged over the active growing period (3–6 months before our collection in this region[37]), reflecting nitrogen availability during this time[38]. This time period spans the main rainy season in Seychelles (December to February), so the nitrogen in these algal tissues should reflect land-based nutrient inputs. Although temporal variability in nutrient input is likely, due to differences in rainfall among years, the spatial patterns associated with adjacent land-based input among sites (that is, the relative differences in nutrient regimes we are interested in here) are thought to be stable through time[38,39].

Wave exposure was calculated based on the uninterrupted (by land or other reefs) distance winds can blow over the ocean to generate waves (fetch), coupled with data on wind speed and direction[40,41]. Larger waves develop with greater fetch and stronger winds. We added reef crest polygons to a base map of land mass[42], and rasterised the map at a spatial resolution of 55 m. Fetch values for each of our 21 reefs was calculated in 32 compass directions (each with an angular width of 11.25) using the Waves Toolbox for ArcGIS 10.2 (ref. 43). We restricted fetch calculations to a maximum distance of 500km around the Seychelles islands, reflecting wind generated wave energy for this isolated archipelago. We used hourly readings of both wind speed and direction from the Seychelles National Meteorological Service for every day of the post-bleaching period (1998 to 2011). Wave energy (in Joules) was calculated as a function of fetch, wind speed and direction[40,41]. We calculated average wave exposure for each reef based on hourly wave exposure estimates for the entire post-bleaching period, capturing information on strong and sporadic winds.

Benthic trajectory. Changes in benthic composition among sites and through time were examined using correlation-based principle components analysis, based on Euclidian distance. Data were log($x + 1$) transformed to account for some right skewness detected in draftsman's plots and normalized to ensure all metrics in the analysis were on a common scale. Eigenvectors were overlaid to identify direction and

contribution of the different variables to the patterns. We defined a regime shifting reef as one where post-disturbance macroalgal cover becomes greater than coral cover and trajectories through time indicate that cover of macroalgae remains high or is increasing. Recovering reefs, conversely, are defined as those reefs where post-disturbance coral cover becomes greater than macroalgal cover and trajectories through time indicate that cover of hard coral remains high or is increasing. To determine whether sites were regime shifting or recovering we used four metrics, the first based on a static cover estimate, while the other three reflect cover trajectories. These four metrics provide a comprehensive assessment of site status and trajectory. Sites must conform to metric 1 and at least one of the trajectory metrics (2–4) to be classified as either recovering or regime shifting.

(Metric 1) Percent cover of coral and macroalgae at last data point (2011). If coral cover is higher than macroalgal cover, and greater than the first post-disturbance survey (2005), the site is classified as recovering. If macroalgal cover is higher than coral cover and greater than pre-disturbance macroalgal cover (1994), the site is classified as regime shifting (Extended Data Table 1).

(Metric 2) The Euclidian distance between the pre-disturbance (1994) benthic composition, and that of 2005, 2008 and 2011 was calculated at a site level, to quantify if this distance was increasing or decreasing through time. A substantial change from 1994 and increasing distance through post-disturbance years indicates a regime shifting site, whereas declining distance through post-disturbance years is indicative of recovery (Fig. 1b, c, Extended Data Fig. 3).

(Metric 3) The rate of change in coral or macroalgal cover post-disturbance (2005–2011) indicates the trajectory of the site in terms of increasing coral versus macroalgae. If rate of coral cover increase remains stable or increases faster than rate of macroalgal cover the site is classified as recovering, and vice versa for regime shifting (Extended Data Table 1).

(Metric 4) Change in cover (coral and macroalgae) between 1994 and 2005, and 1994 and 2011 was calculated. If the net decline in coral cover becomes smaller between these two time periods, and change in macroalgae is negligible and static, the site is classified as recovering. If the net decline in coral cover remains large and static, and increases in macroalgal cover are becoming greater, the site is classified as regime shifting (Extended Data Fig. 4).

To visualize the regime shift and recovery dynamics, we quantified the site level relationships between distance from the pre-disturbance benthic composition and proportions of coral and macroalgae using a multinomial model of benthic composition (j), whereby percentages of coral, macroalgae, and other substrates (100 × π_{ij}) at a given site (i) were predicted based on Euclidian distances (ED) for a given year:

$$\pi_{ij} = \frac{e^{\beta_{0j} + \beta_{1j}ED}}{\sum_{k=1}^{J} (e^{\beta_{0k} + \beta_{1k}ED})} \tag{1}$$

Parameters include an intercept (β_0) and slope (β_1) for ED in each of the habitat categories (j). This is a standard multinomial logit model, in this case run using the *nnet* package in R (http://www.R-project.org). To illustrate the progression of reef states, we ran the same model (1) for each post-disturbance survey (2005, 2008, and 2001) and represented uncertainty in model fit by sampling values from the maximum likelihood and estimated standard deviation of each parameter in (1) and plotting the resulting model fits (Fig. 1c and Extended Data Fig. 3). To better convey the progression of recovering and regime shifting reefs through time we rotated the ED predictor onto the y axis.

Fish functional structure. The functional structure of fish assemblages was calculated based on two dominant traits that capture a large proportion of the implicit functional roles played by reef fishes: dietary group and body size. We classified the fish into well-established feeding groups (corallivores, invertivores, planktivores, grazing herbivores, scraping and excavating herbivores, and piscivores) based on the literature[35] and Fishbase[34]. These groupings cover some of the main feeding functions performed by fishes on coral reefs, including mediation of coral:algae interactions, removal of sediments, and trophic control through predation. Body size is a trait that captures information related to the feeding ecology, energetic demands, and movement patterns in reef fish. For example, there are well established non-linear relationships between body size and area of reef grazed by parrotfishes[22], while predators feed on prey in accordance to their gape size[7]. Furthermore, home range size and functional range size in reef fish is tightly correlated to body size, with larger fish feeding over a greater area[44,45]. Body size was coded into four categories: <20 cm, 20–40 cm, 41–60 cm and >60 cm. We focused on these two biological traits to specifically target functional traits[46].

Based on these two functional traits, we performed a principal coordinate analysis (PCoA) on a Gower distance matrix between species pairs to provide two independent synthetic axes that summarize species distribution within a trait functional space[47,48]. These two independent functional axes from PCoA, in combination with the species abundance matrix for all reefs in each year of sampling, were used to measure functional structure through biomass-weighted mean values for each

community. We used these biomass-weighted mean values for each PCoA axis to assess how the functional structure of a given assemblage changed through time for both recovering and regime shifting reefs. Year and trajectory (regime shift versus recovery) had a significant influence on the functional structure of fish assemblages on both PCoA axes (ANOVA model: axis 1 year $P < 0.001$, trajectory $P < 0.01$; axis 2 year $P < 0.001$, trajectory $P < 0.05$), whereas management (marine reserve versus fished) had no significant influence (ANOVA model: management axis 1 $P = 0.533$, axis 2 $P = 0.106$).

Predicting recovery versus regime shift dynamics. We assessed the ability of eleven different site level factors (habitat type, initial branching coral cover, juvenile coral density, depth, herbivore biomass, herbivore diversity, marine reserve status, nutrient regime, initial structural complexity, sea urchin density, wave exposure) to predict whether a site went on a recovery or regime shift trajectory. Each factor was a priori chosen based on a pre-existing rationale from the literature, to specifically target factors known to represent key processes or to structure coral reef ecosystems globally (Extended Data Table 2). Indeed, the generic importance of the selected variables on coral reefs is highlighted by specific review papers, for example: branching corals[49]; juvenile coral survivorship[2]; depth[20]; marine reserve status[50]; herbivorous fish biomass and diversity[21]; nutrient regime[51]; structural complexity[17]; and sea urchins[2]. Furthermore, the range of values found on Seychelles are consistent with values from other reef locations in the Indo-Pacific (Extended Data Table 3), indicating that the relationships we find should be generic, although we might expect location-specific variation in thresholds. We could not include some factors, such as turbidity and water currents (typically weak in Seychelles), however we were able to accurately predict ecosystem trajectory with the covariates included in the model.

Although some of the factors are not influenced over the time frames investigated here (for example, depth), or were constructed across time (for example, wave exposure), we had to make decisions on which time period to use for other factors. Branching coral cover, important for a range of reef fishes and other organisms[49], is particularly vulnerable to coral bleaching[52], so reefs with a high branching coral cover pre-disturbance may be expected to undergo extensive degradation, and thus we used data from 1994. We used pre-disturbance data for structural complexity because it can be maintained for many years post disturbance, ensuring the continuation of reef processes[13]. For both herbivorous fish biomass and diversity, we used data from 2005 to first account for any changes in these variables associated with the disturbance event itself[53], and second to reflect the amount of post-disturbance herbivory available to influence subsequent benthic responses. We were interested in post-settlement survival based on the juvenile coral density variable, so used data from 2011 to allow more (often slow growing) corals to make it through early life history stages. Sea urchin density (family: Diadematidae), would ideally be taken from 2005 to reflect herbivory available for post-disturbance benthic responses, however the first year of available data was 2008, which correlated well (coefficient 0.57) to subsequent years, suggesting spatial consistency through time. As noted above, spatial differences in our nutrient regime data should reflect long-term patterns. Collinearity among these predictor covariates was assessed (Extended Data Fig. 6), and herbivore diversity was excluded from the analysis due to collinearity with herbivore biomass.

The spatial distribution of sites following a regime shift or recovery trajectory was to some extent geographically clustered, but some sites in close proximity followed different trajectories (Extended Data Fig. 7). We used a Bayesian hierarchical logistic regression model to assess which site-scale variables were best at predicting the trajectory of the sites post bleaching (0 = recovered, 1 = regime shifted). Discarding variables with posterior densities centred on zero (suggesting no effect), our full model for regime shifts among Seychelles sites was:

$$y_{ij} \sim Bern(p_{ij}) \tag{2}$$
$$logit(p_{ij}) \sim N(\mu_j, \sigma) \tag{3}$$
$$\mu_j = \beta_0 + \beta_1 LHB_j + \beta_2 JCD_j + \beta_3 DEP_j + \beta_4 IST_j + \beta_5 CNR_j \tag{4}$$
$$\beta_{0,...,5} \sim N(0.0, 1000) \tag{5}$$
$$\sigma \sim U(0, 100) \tag{6}$$

which included log(herbivore biomass + 1) (LHB; kg ha^{-1}); juvenile coral density (JCD; m^2); depth (DEP; m); initial structural complexity (IST; 0–5 scale); and the carbon nitrogen ratio of sampled algae (CNR). Covariates were then standardized by subtracting their mean and dividing by 2 of their standard deviations to gauge their relative importance as the relative magnitude of their effect sizes. Models were run using the PyMC package[54] for the Python programming language (http://www.python.org).

For each of these variables, we calculated marginal 95% uncertainty intervals (UI) around the point at which the probability of a regime shift was half ($P(shift) = 0.5$). The lower 95% UI was taken as the point at which a regime shift is more likely than not, with all other variables held at their average, while the upper 95% UI was taken

as the point at which recovery is more likely to occur. The values within these uncertainty bounds reflect the unstable area where recovery or regime shift outcomes are equally likely.

In order to assess the predictive ability of the full model, we ran a leave-one-out cross validation (Loo-CV) analysis, a permutation procedure whereby the ability of the model to successfully predict the ecosystem trajectory (recovery or regime shift) of any given site is repeatedly tested with different sites. To do this we randomly selected one site and predicted its recovery or regime shift status using a model based on the other 20 sites, repeating the procedure 999 times.

We tested the predictive power of the model if reduced to include only two metrics (initial complexity of the reef and water depth), both of which can be easily quantified in surveys and collected on reefs before major disturbance events, making them very strong candidates for management planning and informing development of adaptation strategies. This gave a reduced parameter model:

$$y_{ij} \sim Bern(p_{ij}) \tag{7}$$
$$logit(p_{ij}) \sim N(\mu_j, \sigma) \tag{8}$$
$$\mu_j = \beta_0 + \beta_3 DEP_j + \beta_4 IST_j \tag{9}$$

We re-ran the Loo-CV analysis with this subset of predictor variables to assess the predictive power that was lost by removing other variables from the analysis. Despite the substantial increase in deviance information criterion (DIC) scores between models (Mfull DIC = 129; Msubset DIC = 145), the subset model performed equally well in correctly predicting the trajectory of our observed reefs. We sampled from the posteriors of the two variables from the subset model across their observed range to model probability surfaces of regime shift risk when these two metrics are both quantified, presented as a contour plot in Fig. 4.

To assess whether these two simple predictor variables (depth and initial structural complexity) are likely to be useful predictors of post-disturbance reef condition elsewhere in the Indo-Pacific, we compiled data from 6 countries (Australia (Great Barrier Reef), British Indian Ocean Territory (Chagos), Fiji, Kenya, Maldives, Tanzania)[55–58] where disturbance site depth and structural complexity, and post-disturbance (average 8.4 years ± 0.4 post-disturbance) coral and macroalgal cover were recorded. The disturbances were mostly climate induced bleaching, but in combination with crown-of-thorns starfish (COTS) in Fiji and COTS and storms in Australia. For depth we used data from all 51 reef sites. However, because structural complexity can vary based on reef zone (for example, the reef slope versus the reef flat), or geomorphology (for example, atolls versus coastal fringing reefs), we used a paired design for assessing the role of structural complexity where a site that had become dominated by macroalgae was paired with a nearby similar site (for example, same zone, geomorphology, and depth) that had recovered to high coral cover. This resulted in 7 pairs ($n = 14$), from 4 of the countries. We ran the Bayesian hierarchical logistic regression model used for the Seychelles data, with reef sites assigned as having higher macroalgal cover (1) or coral cover (0) post disturbance. The Bayesian effect size posterior density distributions for these wider Indo-Pacific data were compared to those for the Seychelles reefs (Extended Data Fig. 5). Considerable overlap in both the depth (Seychelles: −4.77 [−6.92, −2.62]; Indo-Pacific: −1.67 [−3.74, 0.15]; median and 95% highest posterior density uncertainty intervals) and complexity (Seychelles: −5.66 [−9.13, −2.69]; Indo-Pacific: −3.29 [−6.94, −0.25]) posterior distributions from the six Indo-Pacific countries with those from Seychelles, and more than 95% of the density in the posterior distributions falling below zero in all cases, suggests that both variables are important, generic predictors of macroalgal dominance post-disturbance on Indo-Pacific reefs. Interestingly, although the structure of reefs may vary based on zone or geomorphology, in 6 of the 7 pairwise comparisons, the reef with the greatest structure recovered. On average, reefs that recovered had complexity scores 0.85 (0.09–1.62 UI) greater than reefs than became dominated by macroalgae.

31. Wilson, S. K., Graham, N. A. J. & Polunin, N. V. C. Appraisal of visual assessments of habitat complexity and benthic composition on coral reefs. *Mar. Biol.* **151,** 1069–1076 (2007).
32. Graham, N. A. J. *et al.* Lag effects in the impacts of mass coral bleaching on coral reef fish, fisheries, and ecosystems. *Conserv. Biol.* **21,** 1291–1300 (2007).
33. Letourneur, Y., Kulbicki, M. & Labrosse, P. Length–weight relationship of fishes from coral reefs and lagoons of New Caledonia: an update. *Naga, the ICLARM Quarterly* **21,** 39–46 (1998).
34. Froese, R. & Pauly, D. FishBase (http://www.fishbase.org) (2011).
35. Wilson, S. K. *et al.* Exploitation and habitat degradation as agents of change within coral reef fish communities. *Glob. Change Biol.* **14,** 2796–2809 (2008).
36. Burkepile, D. E. *et al.* Nutrient supply from fishes facilitates macroalgae and suppresses corals in a Caribbean coral reef ecosystem. *Sci. Rep.* **3,** 1493 (2013).
37. Ateweberhan, M., Bruggemann, J. H. & Breeman, A. M. Seasonal dynamics of *Sargassum ilicifolium* (Phaeophyta) on a shallow reef flat in the southern Red Sea (Eritrea). *Mar. Ecol. Prog. Ser.* **292,** 159–171 (2005).

38. Umezawa, Y., Miyajima, T., Yamamuro, M., Kayanne, H. & Koike, I. Fine-scale mapping of land-derived nitrogen in coral reefs by δ^{15}N in macroalgae. *Limnol. Oceanogr.* **47,** 1405–1416 (2002).
39. Wooldridge, S., Brodie, J. & Furnas, M. Expoure of inner-shelf reefs to nutrient enriched runoff entering the Great Barrier Reef Lagoon: Post-European changes and the design of water quality targets. *Mar. Pollut. Bull.* **52,** 1467–1479 (2006).
40. Ekebom, J., Laihonen, P. & Suominen, T. A GIS-based step-wise procedure for assessing physical exposure in fragmented archipelagos. *Estuar. Coast. Shelf Sci.* **57,** 887–898 (2003).
41. Chollett, I. & Mumby, P. J. Predicting the distribution of Montastrea reefs using wave exposure. *Coral Reefs* **31,** 493–503 (2012).
42. Wessel, P. & Smith, W. H. F. GSHHG – A Global Self-consistent, Hierarchical, High-resolution Geography Database. http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html (2014).
43. Rohweder, J. et al. Application of wind fetch and wave models for habitat rehabilitation and enhancement projects – 2012 update. http://www.umesc.usgs.gov/management/dss/wind_fetch_wave_models_2012update.html (2012).
44. Kramer, D. L. & Chapman, M. R. Implications of fish home range size and relocation for marine reserve function. *Environ. Biol. Fishes* **55,** 65–79 (1999).
45. Nash, K. L., Graham, N. A. J. & Bellwood, D. R. Fish foraging patterns, vulnerability to fishing, and implications for the management of ecosystem function across scales. *Ecol. Appl.* **23,** 1632–1644 (2013).
46. McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21,** 178–185 (2006).
47. Legendre, P. & Legendre, L. *Numerical Ecology* 2nd edn (Elsevier, 1998).
48. Villéger, S., Mason, N. W. H. & Mouillot, D. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* **89,** 2290–2301 (2008).
49. Coker, D., Wilson, S. K. & Pratchett, M. S. Importance of live coral habitat for reef fishes. *Rev. Fish Biol. Fish.* **24,** 89–126 (2014).
50. Lester, S. E. et al. Biological effects within no-take marine reserves: a global synthesis. *Mar. Ecol. Prog. Ser.* **384,** 33–46 (2009).
51. Fabricius, K. E. Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Mar. Pollut. Bull.* **50,** 125–146 (2005).
52. Loya, Y. et al. Coral bleaching: the winners and the losers. *Ecol. Lett.* **4,** 122–131 (2001).
53. Pratchett, M. S. et al. Effects of climate-induced coral bleaching on coral-reef fishes: ecological and economic consequences. *Oceanogr. Mar. Biol. Annu. Rev.* **46,** 251–296 (2008).
54. Patil, A. et al. PyMC: Bayesian stochastic modelling in Python. *J. Stat. Softw.* **35,** 1–81 (2010).
55. Cheal, A. J., Emslie, M., MacNeil, M. A., Miller, I. & Sweatman, H. Spatial variation in the functional characteristics of herbivorous fish communities and the resilience of coral reefs. *Ecol. Appl.* **23,** 174–188 (2013).
56. Graham, N. A. J. et al. Climate warming, marine protected areas and the ocean-scale integrity of coral reef ecosystems. *PLoS ONE* **3,** e3039 (2008).
57. Wilson, S. K. et al. Exploitation and habitat degradation as agents of change within coral reef fish communities. *Glob. Change Biol.* **14,** 2796–2809 (2008).
58. McClanahan, T. R., Ateweberhan, M., Darling, E. S., Graham, N. A. J. & Muthiga, N. A. Biogeography and change among regional coral communities across the western Indian Ocean. *PLoS ONE* **9,** e93385 (2014).
59. Jennings, S., Grandcourt, E. M. & Polunin, N. V. C. The effects of fishing on the diversity, biomass and trophic structure of Seychelles' reef fish communities. *Coral Reefs* **14,** 225–235 (1995).
60. Wilson, S. K. et al. Effect of macroalgal expansion and marine protected areas on coral recovery following a climatic disturbance. *Conserv. Biol.* **26,** 995–1004 (2012).
61. Sheppard, C. R. C., Spalding, M., Bradshaw, C. & Wilson, S. Erosion vs. recovery of coral reefs after 1998 El Niño: Chagos reefs, Indian Ocean. *Ambio* **31,** 40–48 (2002).
62. Mumby, P. J. et al. Empirical relationships among resilience indicators on Micronesian reefs. *Coral Reefs* **32,** 213–226 (2013).
63. Cheal, A. J. et al. Coral-macroalgal phase shifts or reef resilience: links with diversity and functional roles of herbivorous fishes on the Great Barrier Reef. *Coral Reefs* **29,** 1005–1015 (2010).
64. Burkepile, D. E. & Hay, M. E. Herbivore species richness and feeding complementarity affect community structure and function on a coral reef. *Proc. Natl Acad. Sci. USA* **105,** 16201–16206 (2008).
65. Rasher, D. B., Hoey, A. S. & Hay, M. E. Consumer diversity interacts with prey defenses to drive ecosystem function. *Ecology* **94,** 1347–1358 (2013).
66. Mumby, P. J. & Harborne, A. R. Marine reserves enhance the recovery of corals on Caribbean reefs. *PLoS ONE* **5,** e8657 (2010).
67. Burkepile, D. E. & Hay, M. E. Herbivore vs. nutrient control of marine primary producers: context-dependent effects. *Ecology* **87,** 3128–3139 (2006).
68. Vergés, A., Vanderklift, M. A., Doropoulos, C. & Hyndes, G. A. Spatial patterns in herbivory on a coral reef are influenced by structural complexity but not by algal traits. *PLoS ONE* **6,** e17115 (2011).
69. Carpenter, R. C. & Edmunds, P. J. Local and regional scale recovery of Diadema promotes recruitment of scleractinian corals. *Ecol. Lett.* **9,** 271–280 (2006).
70. Madin, J. S. & Connolly, S. R. Ecological consequences of major hydrodynamic disturbances on coral reefs. *Nature* **444,** 477–480 (2006).
71. Done, T. J. in *Perspectives on Coral Reefs* (eds. Barnes, D. J. & Clouston, B.) Coral zonation, its nature and significance. 107–147 (Australian Institute of Marine Science, 1983).
72. Larned, S. T. & Atkinson, M. J. Effects of water velocity on NH_4 and PO_4 uptake and nutrient-limited growth in the macroalga *Dictyosphaeria cavernosa*. *Mar. Ecol. Prog. Ser.* **157,** 295–302 (1997).
73. Graham, N. A. J. & McClanahan, T. R. The last call for marine wilderness? *Bioscience* **63,** 397–402 (2013).
74. Graham, N. A. J., Chong-Seng, K. M., Huchery, C., Januchowski-Hartley, F. A. & Nash, K. L. Coral reef community composition in the context of disturbance history on the Great Barrier Reef, Australia. *PLoS ONE* **9,** e101204 (2014).
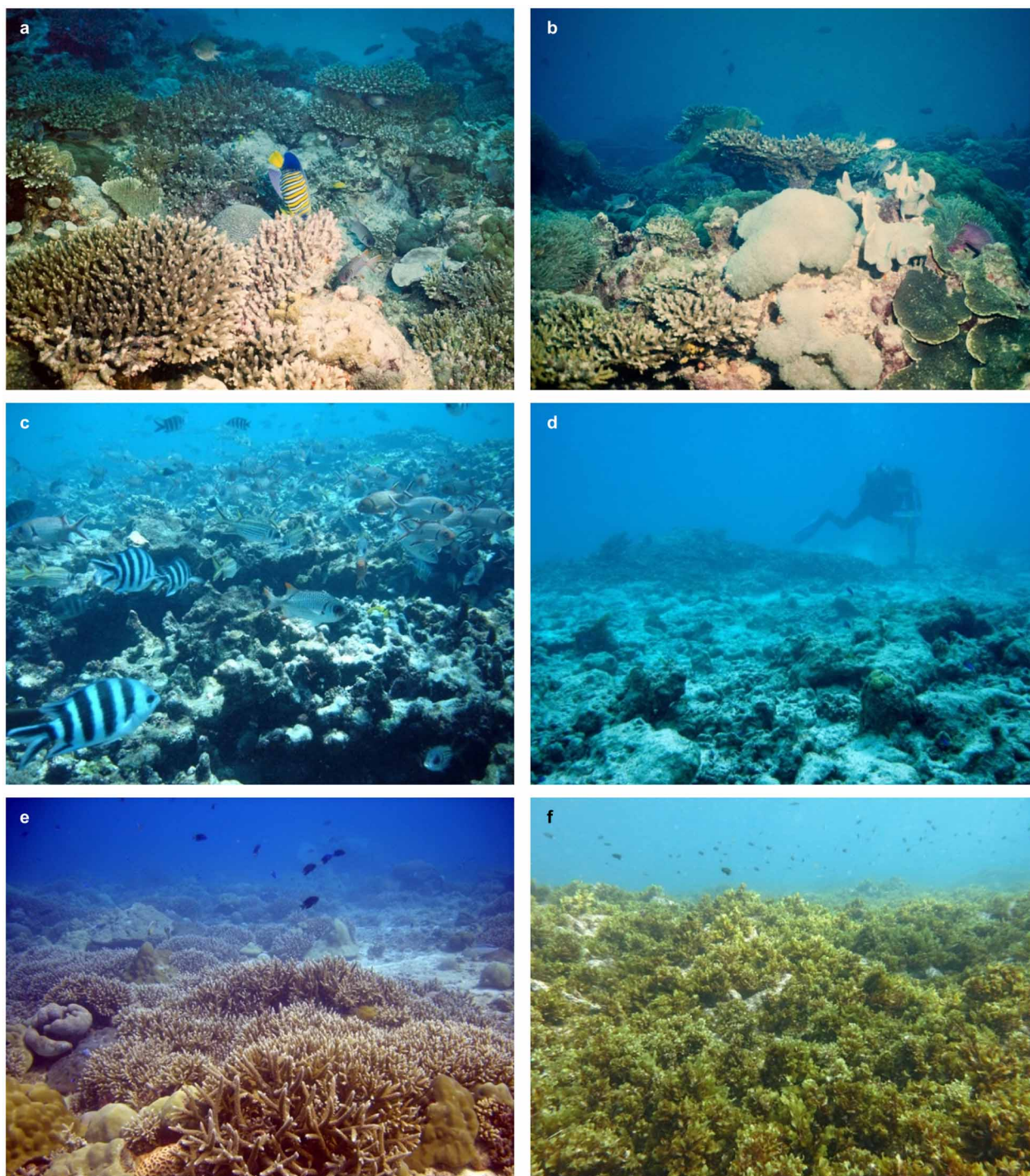
Extended Data Figure 1 | Changing condition of Seychelles coral reefs.
a, b, Coral reefs of the inner Seychelles were typified by high coral cover and low macroalgal cover in 1994. **c, d**, The 1998 coral bleaching event caused widespread coral loss, but some reefs maintained their structural complexity (**c**), while others collapsed (**d**) by 2005. **e, f**, In 2011, many reefs had recovered to high live coral cover (**e**), while others had undergone a regime shift to abundant macroalgal cover (**f**).
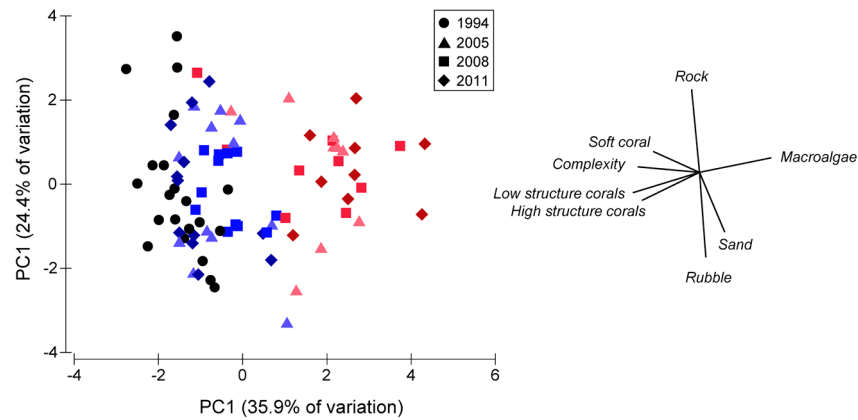
Extended Data Figure 2 | Principal components analysis of benthic composition on 21 reefs across the inner Seychelles 1994–2011. Reefs coloured blue are tracking back to pre-disturbance benthic composition in 2005–2011 following the 1998 bleaching event, whereas reefs coloured red are shifting to alternate benthic compositions, dominated by macroalgae ($n = 84$).
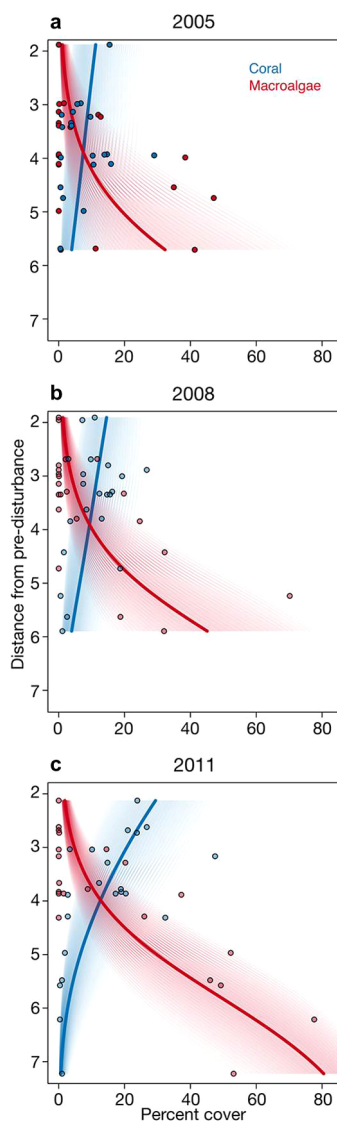
Extended Data Figure 3 | Distance from pre-disturbance benthic community composition. Euclidian distance in multivariate space, plotted against percent cover of dominant biotic benthic organisms (live coral in blue, macroalgae in red) ($n = 63$). **a,** 2005 data. **b**, 2008 data. **c**, 2011 data. Shading represents 95% confidence bounds for the mean trend lines of each habitat type.
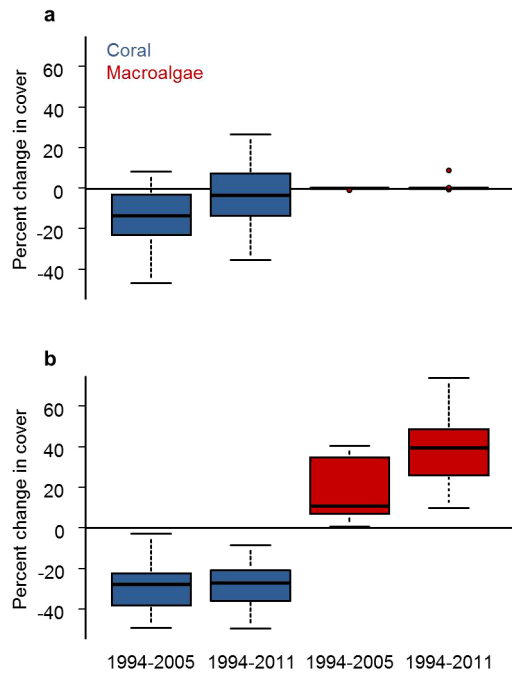
Extended Data Figure 4 | Changing coral and macroalgal cover is relation to pre-disturbance values. a, Data for recovering reefs, where the change in coral cover compared to 1994 was reducing through time, whereas change in macroalgae remained stable ($n = 42$). **b**, Data for regime shifting reefs where the decline in coral cover persisted through time, and changes in macroalgae increased through time ($n = 42$).
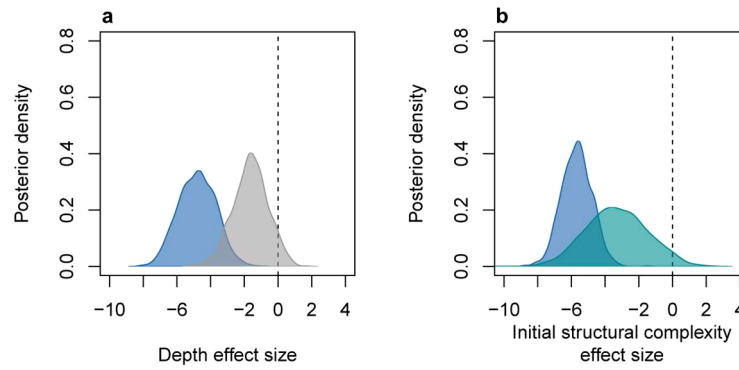
Extended Data Figure 5 | Comparison of effect size posterior density distributions for depth and initial structural complexity in predicting coral versus macroalgae outcomes post disturbance in Seychelles versus 6 other countries across the Indo-Pacific. a, Depth effect size plot, dark blue posterior distribution for Seychelles, grey for other countries ($n = 51$). **b**, Initial structural complexity effect size plot, dark blue distribution for Seychelles, green for other countries ($n = 14$). Depth and structural complexity variables were standardized in both analyses before estimation and all posterior distributions have more than 95% of their density below zero.
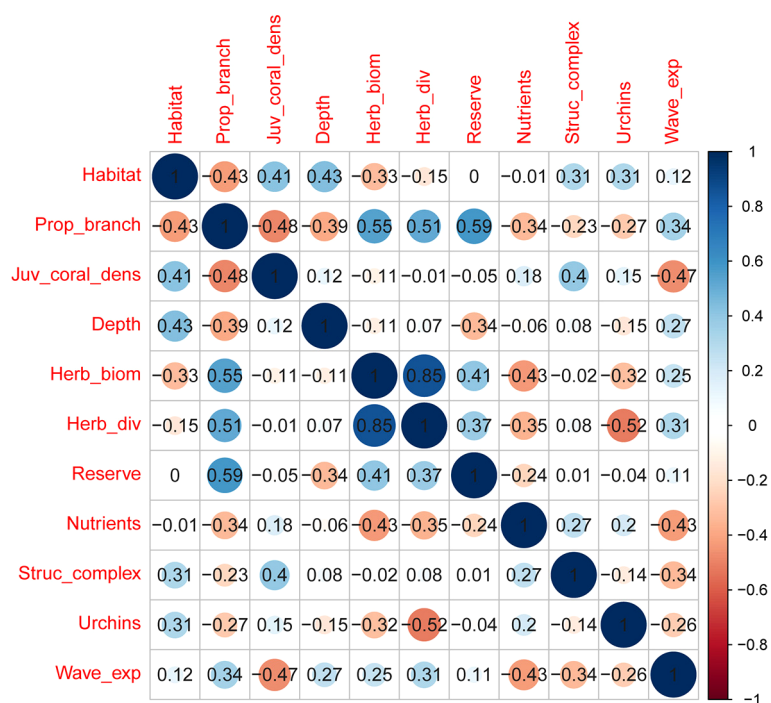
Extended Data Figure 6 | Collinearity matrix of the eleven predictor variables.
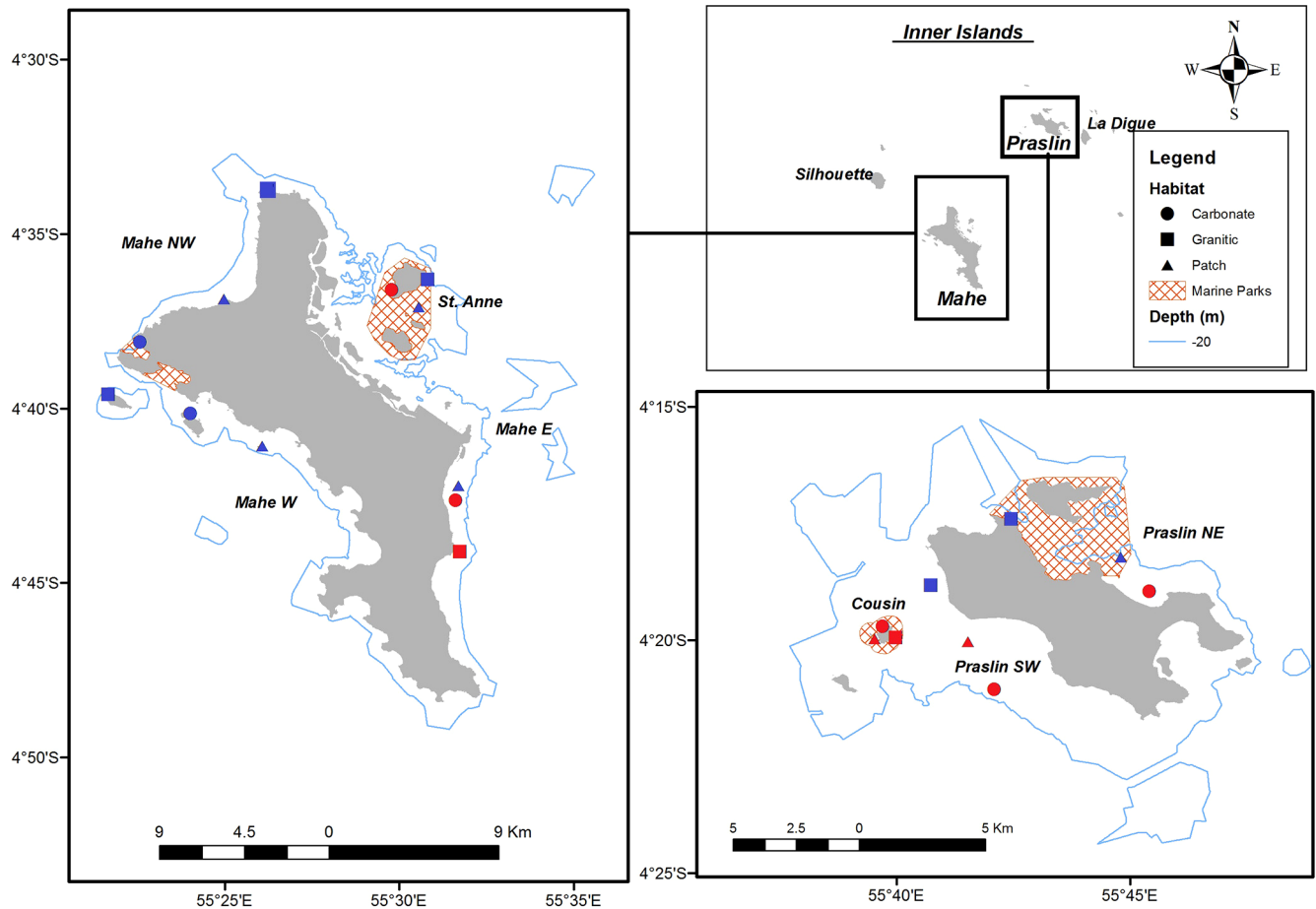
Extended Data Figure 7 | Map of study sites around the inner Seychelles. Sites in blue are recovering from the 1998 mass bleaching event, whereas sites in red have undergone a regime shift to macroalgal cover.

Extended Data Table 1 | Coral cover and macroalgal cover at recovering and regime shifting sites in 2011, with the change (based on the model slope) in coral cover and macroalgae estimates from sites between 2005–2011

	Coral cover (%) 1994	Coral cover (%) 2011	Coral slope 2005-2011	Macroalgae cover (%) 1994	Macroalgae cover (%) 2011	Macroalgae slope 2005-2011
Recovering sites						
Mahe E Patch	19.3	12.3	3.3	0.6	0.1	-0.8
Mahe NW Carbonate	38.9	26.8	8.2	0.0	0.0	0.0
Mahe NW Granitic	10.8	19.0	4.2	0.0	0.0	0.0
Mahe NW Patch	17.3	21.3	3.3	0.0	0.0	0.0
Mahe W Carbonate	34.9	23.9	4.2	0.0	0.0	0.0
Mahe W Granitic	19.3	32.4	8.3	0.0	0.0	0.0
Mahe W Patch	20.8	47.5	9.3	0.0	0.0	0.0
Praslin NE Granitic	17.1	23.8	10.0	0.0	0.1	0.1
Praslin NE Patch	25.4	10.1	2.9	0.0	0.0	0.0
Praslin SW Granitic	17.4	17.5	6.8	0.3	0.0	0.0
Ste. Anne Granitic	38.5	20.4	3.3	1.0	1.5	0.7
Ste. Anne Patch	54.2	18.9	5.7	0.0	8.8	4.4
Average	**26.1 ± 5.5 CI**	**22.8 ± 5.5 CI**	**5.8 ± 1.6 CI**	**0.2 ± 0.2 CI**	**0.9 ± 1.4 CI**	**0.4 ± 0.8 CI**
Regime shift sites						
Cousin Carbonate	49.7	0.4	-0.1	0.4	49.3	4.0
Cousin Granitic	23.3	2.8	0.8	0.6	37.3	16.8
Cousin Patch	38.6	2.8	1.1	0.0	26.0	-4.5
Mahe E Carbonate	28.6	1.9	0.3	12.4	52.3	2.6
Mahe E Granitic	9.3	1.0	-2.5	0.7	46.0	22.2
Praslin NE Carbonate	28.3	1.0	0.3	0.0	53.1	21.0
Praslin SW Carbonate	42.1	0.4	-0.1	3.4	77.6	19.6
Praslin SW Patch	17.3	3.4	1.3	4.7	14.5	1.3
Ste. Anne Carbonate	40.1	14.8	2.6	2.0	20.3	3.8
Average	**30.8 ± 8.5 CI**	**3.14 ± 2.9 CI**	**0.40 ± 0.9 CI**	**2.7 ± 2.6 CI**	**41.8 ± 12.8 CI**	**9.63 ± 6.6 CI**

Average values for coral cover and macroalgal cover in 1994 also given for recovering versus regime shifting reefs.

Extended Data Table 2 | Rationale for predictor variables included in models determining different post-disturbance reef trajectories on Seychelles reefs

Predictor	Rationale	References
Habitat type	Three habitat types were surveyed in Seychelles, including carbonate fringing reefs, granitic rocky reefs, and patch reef habitats. These habitats differ in their underlying matrix that may influence coral recruit success or likelihood of ecosystem collapse.	32,59,60
Pre-disturbance branching coral cover	Branching corals are particularly vulnerable to coral bleaching events, and once dead the structure they provide erodes fairly rapidly. Therefore, reefs with a high cover of branching coral before a disturbance may be particularly vulnerable to extensive coral loss and a reduction in other organisms.	52,61
Juvenile coral density	Successful settlement, survival and growth of new corals is thought to be key to coral recovery dynamics.	2,4
Depth	Many threats on coral reefs are worst in shallow water, making shallow areas most vulnerable to change. Light penetration in shallow water may also favour rapid growth of fleshy macroalgae.	19,20
Herbivorous fish biomass	Herbivorous fish are key to mediating competition for space among corals and algae. Biomass of these fish is a good proxy for function as the area of a reef grazed by these fish scales with both abundance and body size.	21,22,62
Herbivorous fish diversity	Many types of algae have defences against herbivores, meaning that only certain species of fish can feed on some species of algae. This differential ability of fish species to control algae, can mean that a diversity of herbivorous species is required to provide the feeding complementarity necessary to control macroalgae.	63-65
Marine reserve status	No-take marine reserves are expected to reduce fishing and hence enhance ecosystem processes, and may therefore promote faster rates of coral recovery.	28,66
Nutrient regime	Higher nutrient loads in the waters around reefs can enhance the growth of algae and result in algae outcompeting corals, particularly when space becomes available through coral mortality.	19,67
Pre-disturbance structural complexity	The structural complexity of a reef provides a great deal of the habitat variability for a diverse array of other organisms to inhabit. This in turn contributes to enhance a range of ecological processes, and provides niche space for coral settlement and survival. Structural complexity prior to a disturbance is expected to maintain ecosystem processes through the disturbance.	17,18,68
Sea urchin density	Sea urchins are important herbivores on coral reefs, helping to control algal growth and promote successful coral recruitment and recovery.	2,69
Wave exposure	Wave exposure influences coral distribution patterns, growth forms and colony sizes that will likely affect recovery trajectories. Similarly, algal growth can be enhanced with higher flow rates due to increased exposure to water borne nutrients, but algal dislodgement can occur where wave exposure is strong.	70-72

Including references 59–72.

Extended Data Table 3 | Mean values with 95% confidence intervals for predictor covariates in Seychelles compared to other coral reef locations where similar data for the covariates were available

	Seychelles	95% CI	N	Other locations	95% CI	N	Countries
Herbivore biomass	210.3	59.0	21	258.1	33.8	281	Kenya, Madagascar, Maldives, Mauritius, Mayotte, Reunion, Tanzania
Pre-disturbance coral cover	28.1	5.4	21	29.0	3.9	60	Kenya, Tanzania
Pre-disturbance structural complexity	3.2	0.1	21	3.1	0.2	30	British Indian Ocean Territory (Chagos), Great Barrier Reef (Australia)
Urchin abundance	2.2	0.8	21	4.1	2.0	33	Kenya, Madagascar, Mozambique, Tanzania

Data for regions other than Seychelles were provided directly by the authors of these studies[23,73,74].

LETTER

Transferred interbacterial antagonism genes augment eukaryotic innate immune function

Seemay Chou[1]*, Matthew D. Daugherty[2,3]*, S. Brook Peterson[1], Jacob Biboy[4], Youyun Yang[5], Brandon L. Jutras[6,7], Lillian K. Fritz-Laylin[8], Michael A. Ferrin[1], Brittany N. Harding[1], Christine Jacobs-Wagner[6,7,9,10], X. Frank Yang[5], Waldemar Vollmer[4], Harmit S. Malik[2,3] & Joseph D. Mougous[1]

Horizontal gene transfer allows organisms to rapidly acquire adaptive traits[1]. Although documented instances of horizontal gene transfer from bacteria to eukaryotes remain rare, bacteria represent a rich source of new functions potentially available for co-option[2]. One benefit that genes of bacterial origin could provide to eukaryotes is the capacity to produce antibacterials, which have evolved in prokaryotes as the result of eons of interbacterial competition. The type VI secretion amidase effector (Tae) proteins are potent bacteriocidal enzymes that degrade the cell wall when delivered into competing bacterial cells by the type VI secretion system[3]. Here we show that *tae* genes have been transferred to eukaryotes on at least six occasions, and that the resulting domesticated amidase effector (*dae*) genes have been preserved for hundreds of millions of years through purifying selection. We show that the *dae* genes acquired eukaryotic secretion signals, are expressed within recipient organisms, and encode active antibacterial toxins that possess substrate specificity matching extant Tae proteins of the same lineage. Finally, we show that a *dae* gene in the deer tick *Ixodes scapularis* limits proliferation of *Borrelia burgdorferi*, the aetiologic agent of Lyme disease. Our work demonstrates that a family of horizontally acquired toxins honed to mediate interbacterial antagonism confers previously undescribed antibacterial capacity to eukaryotes. We speculate that the selective pressure imposed by competition between bacteria has produced a reservoir of genes encoding diverse antimicrobial functions that are tailored for co-option by eukaryotic innate immune systems.

Eukaryotes can acquire new functions through the exchange of genetic material with other domains of life[1]. Indeed, bacteria-to-eukarya horizontal gene transfer (HGT) underlies the adaptation and diversification of many microbial eukaryotes, such as algae, choanoflagellates and protozoa[4,5]. The acquisition of bacterial genes by metazoans is rare. Among the transferred genes, many are not expressed and have no known function[6], while others have roles in endosymbiont maintenance[7,8]. Relatively few reports provide evidence of transferred elements that confer traits which are directly beneficial to their metazoan recipients[2]. One recent example is the discovery that phytophagous mites and Lepidoptera species exploit a horizontally acquired bacterial cysteine synthase to feed on plants producing cyanogenic defence compounds[9].

Genes that can independently provide new functionality to a recipient organism are strong candidates for domestication after HGT[6,10]. The Tae proteins are small, single-domain enzymes that can rapidly digest the bacterial cell wall[11]. These proteins comprise four phylogenetically distinct families (Tae1–4) that share no overall sequence homology and display unique specificities against peptidoglycan (PG)[3,12]. In the course of probing *tae* distribution, we made the serendipitous observation that homologues are found in distantly related eukaryotic genomic and expression data sets ranging from unicellular protozoa to multicellular metazoans (Fig. 1a). The genes did not appear to derive from contaminating bacterial DNA; most contain introns and are located in genomic regions flanked by eukaryotic genes (Extended Data Fig. 1). We therefore refer to these eukaryotic loci as domesticated amidase effector (*dae*) genes, and hypothesized that they encode antibacterial toxins horizontally acquired from bacteria. Maximum likelihood and Bayesian phylogenetic analyses revealed that trees of bacterial *tae2–4* families each contained two distinct monophyletic clades of eukaryotic *dae* genes (Fig. 1b and Extended Data Figs 2–4). Thus, we conclude that three of the four known *tae* gene families have been acquired by eukaryotes from diverse bacteria in at least six HGT events (Fig. 1a). Our survey is biased by the status of genome sequencing efforts; therefore, these six instances are probably an underestimate of eukaryotic *tae* acquisitions.

Three of the *dae* genes we found are limited to individual or closely related eukaryotes (Fig. 1a, light green, light blue and dark blue). These could represent recent HGT events, or reflect limited genomic and transcriptomic sampling of related species. The remaining three *dae* genes appear to be the result of ancient HGT events. For instance, we found *dae2* in ten species of ticks and mites (Fig. 1b). This dense sampling, a shared intron between the *dae2* genome sequence of *I. scapularis* and *Metaseiulus occidentalis*, and the fact that the tick and mite *dae2* gene phylogeny closely resembles the established phylogeny of these organisms, lead us to conclude that vertical transmission followed a single HGT event of a bacterial *tae2* gene to the common ancestor of ticks and mites approximately 400 million years (Myr) ago (Figs 1b–d and Extended Data Figs 1, 5a, b)[13]. The complete genome sequence of the Acariform mite *Tetranychus urticae* does not possess *dae2*, indicating that loss of the gene has also occurred. Partial *dae2* sequences in the genomes of two scorpion species and the horseshoe crab share an intron position with *dae2* from ticks and mites, suggesting that *dae2* introduction into arthropods may have occurred as early as 550 Myr ago (Extended Data Fig. 5c). Similarly, dense sampling of *dae4* genes in gastropod and bivalve mollusks, as well as a shared *dae4* intron position across all sampled mollusks and an annelid, dates the origin of *dae4* in these animals to at least 400 Myr ago (Fig. 1a, light red, and Extended Data Figs 1, 4)[14]. Finally, a second *dae4* present in a species of choanoflagellates, sea anemones, acorn worms and lancelets is most parsimoniously explained by a single HGT event followed by vertical inheritance and loss in multiple lineages, dating this *dae4* acquisition to before the base of the metazoan lineage (>800 Myr ago) (Fig. 1a, dark red, and Extended Data Fig. 4). However, owing to sparse sampling and lack of evidence of shared synteny, we cannot rule out more recent HGT to and between these eukaryotic

[1]Department of Microbiology, University of Washington School of Medicine, Seattle, Washington 98195, USA. [2]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [3]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [4]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4AX, UK. [5]Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [6]Microbial Sciences Institute, Yale University, New Haven, Connecticut 06516, USA. [7]Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06516, USA. [8]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA. [9]Department of Microbial Pathogenesis, Yale University, New Haven, Connecticut 06516, USA. [10]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06516, USA.
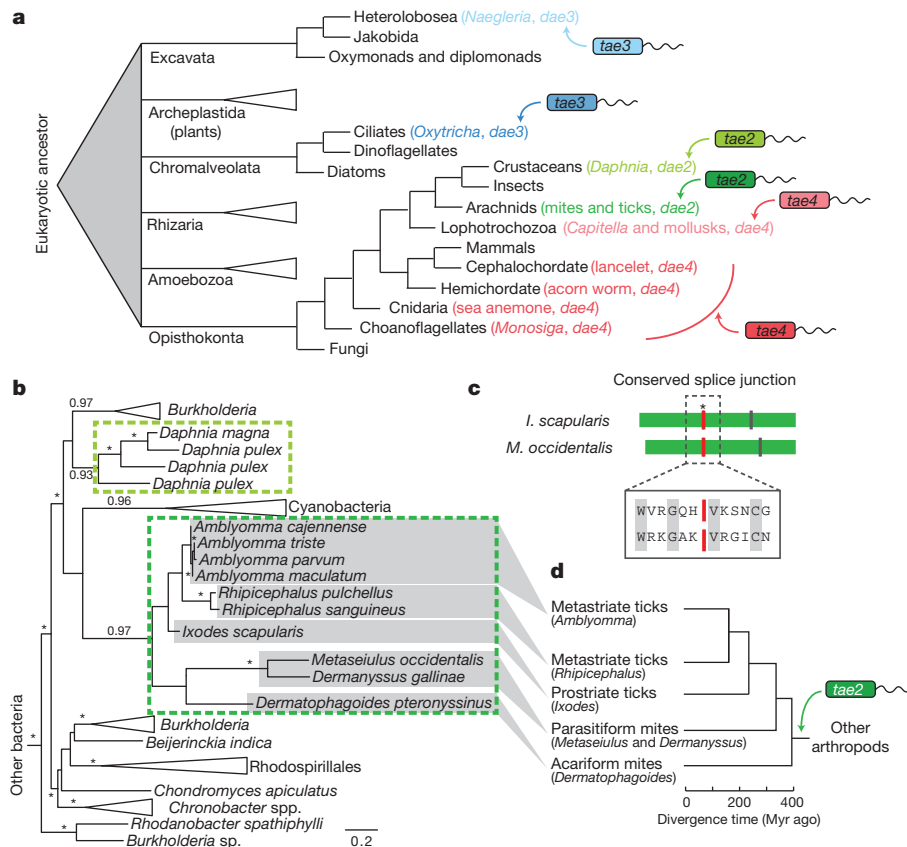*These authors contributed equally to this work.

Figure 1 | Recurrent horizontal gene transfer of *tae* genes into diverse eukaryotic lineages. a, Schematized phylogenetic tree of basal eukaryotic lineages[27] showing instances of *tae* transfer (arrows) from bacteria to eukaryotes, coded by colour (*tae* family) and shading (acquisition events). **b**, Maximum likelihood phylogenetic tree of *tae2* and *dae2* genes. Representatives are boxed and colour-coded according to **a**. Branch support >0.7 indicated by asterisks or numbers. Scale bar shows estimated divergence in amino acid changes per residue. Dashed lines highlight separate HGT events. **c**, Schematic alignment of tick (*I. scapularis*) and mite (*M. occidentalis*) *dae2* genes with shared (red line, asterisk) and unique (vertical lines) intron positions denoted. Aligned residues surrounding the splice site are shown (boxed) with conserved amino acids indicated (grey). **d**, Tick and mite phylogeny with approximate dates of divergence based on concordance with the *dae2* gene tree (**c**)[13].

lineages[4]. In summary, we find compelling evidence that at least two animal lineages have retained a bacterially derived antibacterial gene for hundreds of millions of years.

Several lines of evidence led us to hypothesize that *dae* genes provide an adaptive function to their eukaryotic hosts. We found strong signatures of purifying selection acting on *dae2* and *dae4* genes (Extended Data Table 1). Additionally, eukaryotic Sec signals were identified in the majority of Dae proteins, including representatives from each of the predicted HGT events (Extended Data Fig. 6). Secretion of bacterial Tae proteins occurs through the Sec-independent type VI secretion system (T6SS); thus, acquisition of a Sec signal is indicative of functional specialization involving export from eukaryotic cells. Finally, the majority of Dae proteins possess the cysteine–histidine catalytic dyad and flanking motifs of their corresponding Tae families, consistent with retention of enzymatic activity (Extended Data Fig. 6).

We next sought evidence for expression of eukaryotic *dae* homologues belonging to each of the transferred bacterial *tae* families. We found *dae2* expression during both the unfed nymphal and unfed adult life stages of the hard tick *I. scapularis*, with levels significantly elevated in adults (Fig. 2a). In the amoeba *Naegleria gruberi*, we observed a basal level of expression of each of the three *dae3* homologues in trophozoite (amoeba) cells, which increased during differentiation into flagellates (Fig. 2b). A published expression profile of the lancelet *Branchiostoma floridae* indicates that expression of *dae4* is enriched at the neurula stage of development[15]. Together, these data strongly support the hypothesis that *dae* genes have been functionally integrated into recipient physiology.

The Tae families display unique specificities against PG. Within PG typified by Gram-negative Proteobacteria, enzymes from families 1 and 4 cleave at the γ-D-glutamyl-*meso*-diaminopimelic acid (*m*DAP) bond, whereas those from families 2 and 3 cleave the *m*DAP-D-alanine bond crosslinking the peptide stems (Fig. 2c)[3,12,16]. To test whether Dae proteins can hydrolyse PG, we incubated purified Dae2, Dae3 and Dae4 representatives from *I. scapularis*, *N. gruberi* and *B. floridae*, respectively, with isolated *Escherichia coli* PG sacculi. High-performance liquid chromatography (HPLC) analysis of reaction products demonstrated that each of the enzymes hydrolyses PG (Fig. 2d, e). Remarkably, Dae2, Dae3 and Dae4 show substrate specificity matching that of the characterized extant Tae homologues within corresponding families (Fig. 2c). These data support the hypothesis that *dae* homologues, derived from three *tae* families, have been retained in eukaryotic genomes due to their PG amidase activity. We did not find evidence supporting the transfer of housekeeping bacterial amidases to these organisms, leading us to speculate that genes encoding T6S effectors—enzymes that intoxicate recipient cells at exceedingly low concentrations—might be especially amenable to preservation after HGT[17].

Within eukaryotes, enzymes with PG-degrading activity might have immunoregulatory roles, or act directly as antibacterial factors like the Tae toxins[18]. To explore the functional significance of a domesticated *tae* gene, we focused on *dae2* from the deer tick *I. scapularis*, an important vector for numerous diseases, including Lyme borreliosis and anaplasmosis[19]. Western blot analysis of adult *I. scapularis* demonstrated that Dae2 is present in the salivary glands and midgut (Fig. 3a). *I. scapularis* is an ectoparasite that requires a blood meal for life-stage transitions; pathogens are typically acquired during feeding and transmitted to a new host at the next blood meal. Accordingly, the midgut and salivary glands interface with bacterial pathogens and influence their transmission[20].
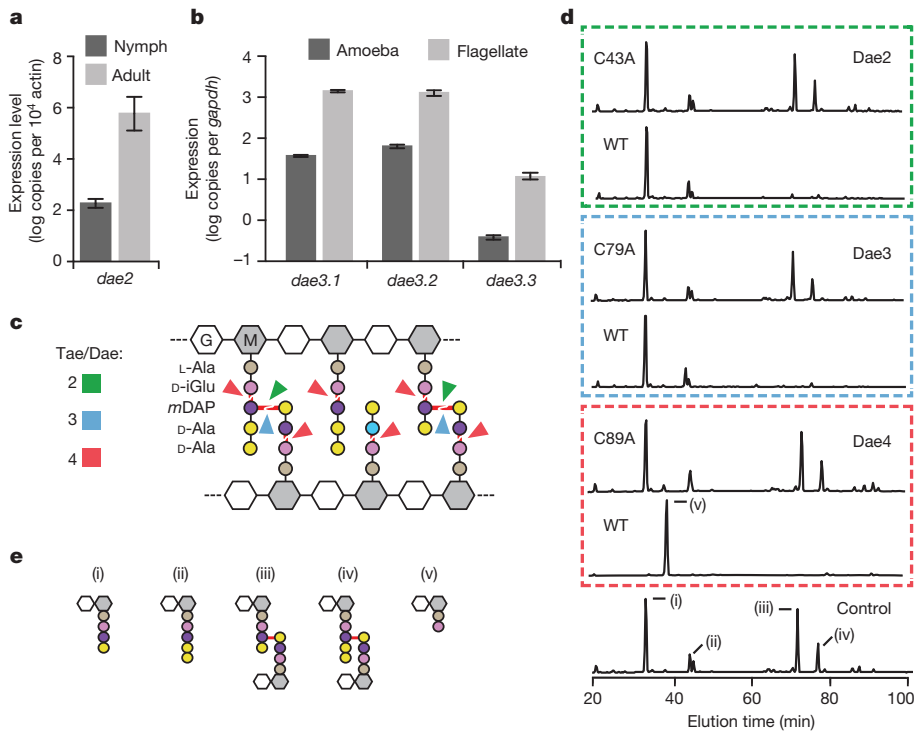
Figure 2 | Eukaryotic *dae* genes encode differentially expressed PG amidases with conserved specificity. a, b, Expression profile of *I. scapularis* and *N. gruberi dae* genes at the indicated life stages as measured by polymerase chain reaction with quantitative reverse transcription (qRT–PCR). Levels of each *N. gruberi dae3* gene (*dae3.1–3.3*) were determined. Error bars show ± standard deviation (s.d.), *n* = 3. **c**, Schematic representation of typical Gram-negative PG showing cleavage sites (red lines) for Tae and Dae families 2–4 (colours correspond to Fig. 1a). **d**, Partial HPLC chromatograms of *E. coli* PG sacculi products resulting from incubation with buffer (control), wild-type (WT) and catalytically inactive (C43A, C79A, C89A) Dae enzymes and cellosyl. **e**, Major HPLC peaks assigned previously by mass spectrometry correspond to disaccharide-linked tetrapeptide (i), pentapeptide (ii), tetrapeptide–tetrapeptide (iii), pentapeptide–tetrapeptide (iv) and dipeptide (v)[3].

To understand how Dae2 could contribute to innate bacterial defence within *I. scapularis*, we tested its capacity to cleave diverse PG structures representative of bacteria the organism encounters in the environment[21]. Consistent with its ability to degrade *E. coli* PG, we found that Dae2 degrades a related form of the cell wall present in Firmicutes belonging to the class Bacilli (Extended Data Fig. 7a)[16]. We did not detect cleavage of the lysine-type PG found in *Streptococcus pneumoniae*, which represents the second major PG type found in Firmicutes (Extended Data Fig. 7b). Although the ultrastructure of the *B. burgdorferi* sacculus is not well defined, its amino acid composition appears to differ from that of

well-characterized bacterial cell walls[22]. Incubation of *B. burgdorferi* sacculi with Dae2 led to the accumulation of specific enzymatic degradation products, indicating that the cell wall of this organism is also a substrate of the amidase (Extended Data Fig. 8).

The Dae proteins are reminiscent of an evolutionarily conserved group of bacteriophage-related eukaryotic innate immune amidases, the PG recognition proteins (PGRPs)[18]. Some PGRPs are directly bacteriocidal and act by hydrolysing PG, whereas others exert antibacterial activity through alternative mechanisms[23]. We found that exogenous Dae2 is not toxic to intact *E. coli* cells. By contrast, Dae2, but not a catalytically inactive variant of the enzyme (C43A), administered to outer-membrane-permeabilized *E. coli* or targeted to the periplasm via the Sec pathway, is highly lytic (Fig. 3b–d). Moreover, exogenous Dae2 is bacteriocidal against *B. subtilis*, which has cell-surface-exposed PG (Fig. 3d). Together, these results strongly suggest that Dae2-dependent antibiosis is solely the result of its amidase activity and that the enzyme would require outer membrane permeabilizing agents such as antimicrobial peptides to act *in vivo*.

B. burgdorferi is the causative agent of Lyme disease, the most prevalent vector-borne illness in the United States[24]. Given the antibacterial

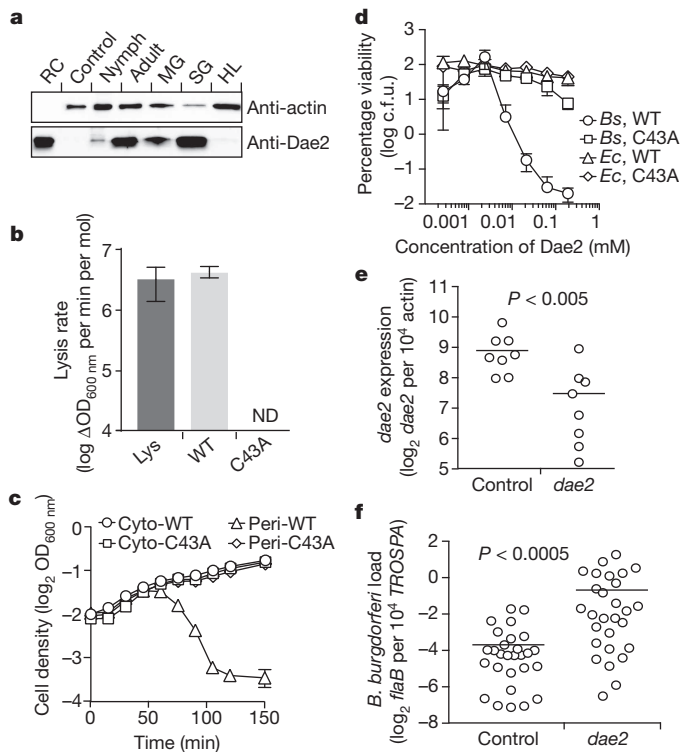**Figure 3 | Dae2 is a bacteriolytic toxin that restricts the proliferation of *B. burgdorferi* in the tick *I. scapularis*. a**, Western blot analysis of Dae2 in unfed adult and nymphal total tissue (total), midgut (MG), salivary gland (SG) and haemolymph (HL) extracts from *I. scapularis*. Recombinant Dae2 protein (RC) and tissue from a closely related species, *Dermacentor variablis* (control), were included. Actin levels were examined as a loading control. **b**, Lytic activity of lysozyme (Lys) and Dae2 (wild type (WT), C43A) proteins against permeabilized *E. coli*. Error bars show ± s.d., *n* = 3. ND, not detected. **c**, Growth of *E. coli* expressing native (cyto-) or periplasm-targeted (peri-) Dae2 proteins. OD$_{600 \, nm}$, optical density at 600 nm. Error bars show ± s.d., *n* = 3. **d**, Bacterial killing activity of indicated proteins against *B. subtilis* (*Bs*) and *E. coli* (*Ec*) cells. c.f.u., colony-forming units. Error bars show ± s.d., *n* = 3. **e**, *Dae2* transcript levels quantified by qRT–PCR in RNAi-treated engorged ticks. **f**, At 2 weeks post-engorgment, spirochaete levels were quantified in ticks that had received the indicated RNAi treatments, using qPCR analysis of *flaB*, a *B. burgdorferi*-specific gene, and normalized to *TROSPA*, a tick-specific gene. *n* = 20. Each data point in **e** and **f** represents three nymphs. Horizontal bars represent mean values, which were significantly different in a two-tailed nonparametric Mann–Whitney test (*P* < 0.05).

activity of Dae2 (Fig. 3b–d), its ability to cleave *B. burgdorferi* PG *in vitro* (Extended Data Fig. 8), and its localization to sites that interface with bacteria (Fig. 3a), we hypothesized that Dae2 could have a role in regulating *B. burgdorferi* populations in *I. scapularis*. We tested this possibility using RNA interference (RNAi)-mediated knockdown of *dae2* (Fig. 3e). RNAi-treated nymphal ticks were fed to repletion on *B. burgdorferi*-infected mice, and spirochaete load was assessed at engorgement and again after 2 weeks. At repletion, we observed no detectable difference in *B. burgdorferi* levels in control and experimental RNAi-treated ticks, indicating that Dae2 activity does not limit initial acquisition of the bacterium (Extended Data Fig. 9a). By contrast, at 2 weeks post-engorgement, *B. burgdorferi* levels were significantly elevated in the *dae2* knockdown group (Fig. 3f). The effect of Dae2 disruption on *B. burgdorferi* levels is unlikely to be due to variations in tick feeding or general fitness, as we observed no difference between the groups in engorgement weights at either time point (Extended Data Fig. 9b). Furthermore, overall bacterial load was similar between the groups, suggesting that the increase in *B. burgdorferi* did not result from gross changes in populations of tick-associated microbes (Extended Data Fig. 9c). The ability of Dae2 to act on a wide range of bacterial cell walls leaves open the possibility that compositional changes to the tick microbiome may contribute to the effect of the knockdown on *B. burgdorferi*[25]. On the basis of these findings, we conclude that Dae2 contributes to the innate ability of *I. scapularis* to control *B. burgdorferi* levels after its acquisition. This has potential ramifications for Lyme disease transmission, as spirochaete load in the tick can influence transmission efficiency[26].

We demonstrate that bacterial genes encoding antibacterial effectors of the T6SS have been horizontally transferred to diverse eukaryotes. The recurrent and independent transfer of *tae* genes to distinct eukaryotic lineages suggests that these toxins can confer immediate fitness benefits by supplying new function to the innate immune system[10]. Recent studies have revealed that the number and diversity of factors that mediate interbacterial antagonism is greater than once appreciated. Thus, we speculate that competition between bacteria generates a reservoir of genes—beyond the *tae* superfamily—with the potential to confer antimicrobial capacity to eukaryotes upon acquisition.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 August; accepted 13 October 2014.
Published online 24 November 2014.

1. Boto, L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc. R. Soc. B* **281,** 20132450 (2014).
2. Dunning Hotopp, J. C. Horizontal gene transfer between bacteria and animals. *Trends Genet.* **27,** 157–163 (2011).
3. Russell, A. B. *et al.* A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe* **11,** 538–549 (2012).
4. Andersson, J. O. Gene transfer and diversification of microbial eukaryotes. *Annu. Rev. Microbiol.* **63,** 177–193 (2009).
5. Schönknecht, G. *et al.* Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339,** 1207–1210 (2013).
6. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nature Rev. Genet.* **9,** 605–618 (2008).
7. Husnik, F. *et al.* Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153,** 1567–1578 (2013).
8. Nikoh, N. & Nakabachi, A. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol.* **7,** 12 (2009).
9. Wybouw, N. *et al.* A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *eLife* **3,** e02365 (2014).
10. Moran, Y., Fredman, D., Szczesny, P., Grynberg, M. & Technau, U. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. *Mol. Biol. Evol.* **29,** 2223–2230 (2012).
11. Russell, A. B. *et al.* Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475,** 343–347 (2011).
12. Chou, S. *et al.* Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the type VI secretion system. *Cell Rep* **1,** 656–664 (2012).
13. Jeyaprakash, A. & Hoy, M. A. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp. Appl. Acarol.* **47,** 1–18 (2009).
14. Warnke, K. M., Meyer, A., Ebner, B. & Lieb, B. Assessing divergence time of Spirulida and Sepiida (Cephalopoda) based on hemocyanin sequences. *Mol. Phylogenet. Evol.* **58,** 390–394 (2011).
15. Yu, J. K. *et al.* A cDNA resource for the cephalochordate amphioxus *Branchiostoma floridae. Dev. Genes Evol.* **218,** 723–727 (2008).
16. Vollmer, W., Blanot, D. & de Pedro, M. A. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32,** 149–167 (2008).
17. Russell, A. B., Peterson, S. B. & Mougous, J. D. Type VI secretion system effectors: poisons with a purpose. *Nature Rev. Microbiol.* **12,** 137–148 (2014).
18. Dziarski, R. & Gupta, D. The peptidoglycan recognition proteins (PGRPs). *Genome Biol.* **7,** 232 (2006).
19. Sonenshine, D. E. & Roe, R. M. *Biology of Ticks* 2nd edn (Oxford Univ. Press, 2013).
20. Hajdušek, O. *et al.* Interaction of the tick immune system with transmitted pathogens. *Front Cell Infect Microbiol* **3,** 26 (2013).
21. Hawlena, H. *et al.* The arthropod, but not the vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. *ISME J.* **7,** 221–223 (2013).
22. Beck, G., Benach, J. L. & Habicht, G. S. Isolation, preliminary chemical characterization, and biological activity of *Borrelia burgdorferi* peptidoglycan. *Biochem. Biophys. Res. Commun.* **167,** 89–95 (1990).
23. Kashyap, D. R. *et al.* Peptidoglycan recognition proteins kill bacteria by activating protein-sensing two-component systems. *Nature Med.* **17,** 676–683 (2011).
24. Radolf, J. D., Caimano, M. J., Stevenson, B. & Hu, L. T. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nature Rev. Microbiol.* **10,** 87–99 (2012).
25. Narasimhan, S. *et al.* Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the lyme disease spirochete. *Cell Host Microbe* **15,** 58–71 (2014).
26. Zhang, L. *et al.* Molecular interactions that enable movement of the Lyme disease agent from the tick gut into the hemolymph. *PLoS Pathog.* **7,** e1002079 (2011).
27. Keeling, P. J. *et al.* The tree of eukaryotes. *Trends Ecol. Evol.* **20,** 670–676 (2005).

Author Contributions S.C., M.D.D., H.S.M. and J.D.M. designed the study. S.C., M.D.D., S.B.P., J.B., Y.Y., B.L.J., L.K.F.-L., M.A.F., B.N.H., C.J.-W., X.F.Y., W.V., H.S.M. and J.D.M. performed experiments, analysed data and provided intellectual input into aspects of this study. S.C., M.D.D., S.B.P., H.S.M. and J.D.M. wrote the manuscript; all authors contributed to its editing.

METHODS

Computational searches. Homologues of *tae* were searched for using iterative PSI-BLAST[29]. First, bacterial homologues were assembled using PSI-BLAST searches limited to bacterial sequences in the non-redundant (NR) protein database. Sequences with *e*-values $< 1 \times 10^{-10}$ (for *tae2*) or $< 1 \times 10^{-20}$ (for *tae1*, *tae3* and *tae4*) and greater than 50% query coverage were included in successive rounds until no new homologues were identified. Position-specific scoring matrix (PSSM) was used to query the NR database limited to eukaryotic sequences. Eukaryotic homologues with *e*-values $< 1 \times 10^{-5}$ were used to initiate an iterative PSI-BLAST search for eukaryotic proteins, as described earlier (*e*-value cut-off 1×10^{-5}). Eukaryotic homologues were validated by the presence of introns or flanking eukaryotic genes. To validate the *Oxytricha trifallax dae3* gene identified in the macronucleus genome, we searched the unpublished micronucleus genome (http://oxy.ciliate.org/blast/) for evidence of a fragmented *dae3* gene that would be consistent with gene rearrangement in this species[28]. Expressed sequence tag (EST), whole-genome sequencing (WGS) and transcriptome databases were searched with tBlastN[29] using validated *dae* genes. We acknowledge the deposition of unpublished data into these databases from multiple sources, including Baylor College of Medicine Human Genome Sequencing Center (https://www.hgsc.bcm.edu), The Genome Institute at Washington University School of Medicine (http://genome.wustl.edu), the US Geological Survey (http://www.usgs.gov), the Functional Genomics Center Zurich (http://www.fgcz.ch), the Joint Genome Institute (http://jgi.doe.gov) and the Broad Institute (http://www.broadinstitute.org). Hits from EST or transcriptome databases were accepted in cases where the hit was more closely related to a validated *dae* gene than a bacterial *tae* gene. When gene predictions based on genomic sequences differed from experimental data from EST or transcriptome databases, we used experimental data to confirm or modify the predicted protein sequence. For instance, the predicted *dae2* gene from *I. scapularis* (NCBI protein database accession gi|242000170) lacks a secretion signal, whereas the sequence from EST data (NCBI EST database accession gi|156264544) differs from the sequence in the protein database in the first exon, resulting in a strongly predicted secretion signal similar to the other tick sequences.

Phylogenetic and evolutionary analysis. Bacterial and eukaryotic sequences were aligned using MUSCLE[30] and edited using Geneious[31]. Regions encompassing the catalytic domain were used for phylogenetic analyses; sequences with $>98\%$ identity were excluded. The best-fitting evolutionary model was determined by Prottest[32]. Maximum likelihood phylogenetic trees were generated with PhyML[33] using 500 bootstrap replicates. To validate phylogenetic inferences, Bayesian Markov chain Monte Carlo (MCMC) analyses were performed in MrBayes[34], sampling every 500 generations until the standard deviation of split frequencies was <0.01 or 10^6 generations were sampled. Tests for purifying selection were performed on aligned and degapped nucleotide sequences of *dae* or *tae* genes. Whole gene non-synonymous/synonymous (dN/dS) ratio calculations, as well as statistical tests for purifying or positive selection for individual codons, were performed using SLAC in the HyPhy software suite[35]. Additional statistical tests in the HyPhy software suite (REL and FUBAR) confirmed that several *tae* and *dae* codons display statistically significant signatures of purifying selection. No codons demonstrate signatures of positive selection. N-terminal eukaryotic secretion signals were predicted using SignalP[36] using default cut-off values. Sequence logos were constructed using Geneious.

DNA libraries. For *N. gruberi* cDNA libraries, strain NEG was grown on *Klebsiella* and differentiated using standard protocols[37]. Synchrony was estimated by percentage of flagellates after fixing in Lugol's iodine ($n > 100$ per time point)[38]. 10^7 cells were harvested per sample. For *I. scapularis* cDNA libraries, ticks from the Tick-Rearing Center at Oklahoma State University were homogenized by grinding in liquid nitrogen. RNA and DNA was purified from *I. scapularis* and *N. gruberi* samples with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Contaminating genomic DNA in RNA samples was removed by treatment with Turbo DNase (Invitrogen) for 1 h at 37 °C, followed by a second TRIzol purification. DNA contamination was checked by PCR using actin- or GAPDH-specific primers for *I. scapularis* and *N. gruberi*, respectively. cDNA libraries were synthesized using the iScript cDNA synthesis kit (Biorad).

Expression of Dae proteins. The codon-optimized *dae* genes from *I. scapularis* (*dae2*), *N. gruberi* (*dae3*) and *B. floridae* (*dae4*) with predicted signal sequences removed were synthesized by GenScript and cloned into the pHis-sumo expression vector. Shuffle T7 pLysS cells were transformed with plasmids, and expression was induced at an optical density ($OD_{600\,nm}$) of 0.6 with 0.1 mM isopropyl-β-D-thiogalactopyranoside for 20 h at 18 °C. Cells resuspended in 20 mM HEPES pH 7.5, 0.5 M NaCl, 25 mM imidazole were lysed by sonication. Lysate was cleared by centrifugation for 1 h at 18,000*g*, and proteins were purified with a metal-chelating affinity column. The tag was proteolytically removed with the H3C and separated from proteins using a second affinity column and size exclusion chromatography (GE Healthcare).

Sacculus analysis. PG sacculi from *E. coli* D456 ($\Delta dacA\ \Delta dacB\ \Delta dacC$) were purified as previously described[39,40]. Preparations (300 μg) were incubated with Dae2 (1 μM),

Dae3 (10 μM) or Dae4 (10 μM) in 300 μl of 20 mM HEPES pH 7.5, 100 mM NaCl for 4 h at 37 °C. PG sacculi from *B. subtilis* 168 (300 μg) or from *S. pneumoniae* R6 (120 μg) were incubated with Dae2 (6 μM) for 4 h at 37 °C. The samples were digested with cellosyl, reduced and analysed by HPLC using published methods[40]. For preparations from *B. burgdorferi*, the B31-MI-16 strain, an infectious clone of the sequenced type strain B31, was cultured at 34 °C to early mid-log exponential growth[41–43]. Cultures were chilled on ice for 10 min and gently harvested by centrifugation at 3,250*g* for 15 min. Pelleted cells were washed three times and resuspended in cold PBS. Cell suspensions were added, drop-wise, to 6 ml of 8% SDS and boiled for 30 min. PG was prepared, incubated with Dae2 (1 μM) for 4 h at 37 °C, and analysed as previously described[40], with the exception of HPLC buffer B, which contained 30% methanol.

Western blot analysis. Tissues were dissected from *I. scapularis* ticks purchased from the Tick-Rearing Center at Oklahoma State University. A rabbit polyclonal antibody specific for *I. scapularis* Dae2 was generated by GenScript using a synthetic peptide corresponding to Dae2 amino acids 123–136 (RYGNTGKPNYNGDN, Lot #195690-4). Mouse anti-actin antibody from Abcam (GR14272-8) and anti-rabbit (A6154) and anti-mouse (A4416) horseradish peroxidase (HRP)-conjugated secondary antibodies from Sigma were used. Western blot analyses and imaging were performed as previously described[44]. Four replicate analyses of tissues were performed; a representative blot is shown in Fig. 3a.

Growth curves. *E. coli* growth curves were generated as previously described[11]. The vector pSCHRAB2 was used for expression of cytoplasmic *I. scapularis* Dae2, and the pSCRHAB2 vector with a *pelB* leader sequence inserted was used for expression of periplasmic Dae2. Curves are representative of three biological experiments and contain technical triplicates.

Lysis assays. Assays were performed as previously described[45]. *E. coli* reactions were carried out at enzyme concentrations of 1 μM; *B. subtilis* reactions were carried out at concentrations of 1 μM (lysozyme) and 50 μM (Dae2). Curves are representative of three biological experiments and contain technical triplicates.

Bacterial killing assays. Colonies of *E. coli* or *B. subtilis* cells grown on solid LB media were washed in 0.2× PBS pH 6 and resuspended to an $OD_{600\,nm}$ of 0.1 and 0.01, respectively. Cells were incubated with recombinant Dae2 enzyme (wild type or C43A) at room temperature for 3 h, and serial dilutions were plated on solid LB. Viability was quantified by enumeration of colony-forming units. Curves contain four technical replicates.

Mouse and RNAi experiments. All animal experiments and tick protocols were approved by the Institutional Animal Care and Use Committee at Indiana University. The low-passage, virulent *B. burgdorferi* strain 5A4NP1, a derivative of B31-MI, was a gift from H. Kawabata and S. Norris, University of Texas Health Science Center. The strain was cultivated in Barbour–Stoenner–Kelly (BSK-II) medium supplemented with 6% normal rabbit serum (Pel Freez Biologicals) at 35 °C with 5% CO_2. Kanamycin was added to the culture at 300 μg ml^{-1}. The mouse feeding experiments were conducted in the Vector-borne Diseases Laboratory at Indiana University School of Medicine. Briefly, 4-week-old C3H/HeN mice were needle-infected with *B. burgdorferi* (10^5 spirochetes per mouse). Two weeks post-inoculation, mouse infection was confirmed by cultivation of ear-punch biopsy specimens to assess spirochaete growth. A single growth-positive culture was used as the criterion for infection of each mouse.

RNAi in nymphal ticks was performed using previously described protocols[46]. To generate double-stranded RNA (dsRNA), 374 bp of *I. scapularis dae2* and 356 bp of the green fluorescent protein gene (*gfp*) were amplified using the following primers containing the T7 promoter: gfp_RNAi_F, GAGCTCTAATACGACTCACTATAGGGAGAGTGTGAGTTATAGTTGTATTCCAAT; gfp_RNAi_R, GGTACCTAATACGACTCACTATAGGGAGAGTGGAGAGGGTGAAGGTGATGCAAC; dae2_RNAi_F, CTAGTCGAGCTCTAATACGACTCACTATAGGGAGACGCTCGTGGTCCTGGGAT; dae2_RNAi_R, CTAGTCGGTACCTAATACGACTCACTATAGGGAGAGGTTGTAGTTGGGCTTCCCTGTA. dsRNA was synthesized and purified from PCR products using a commercial kit (Megascript RNAi Kit, Ambion), and resuspended into elution buffer (10 mM Tris-HCl pH 7, 1 mM EDTA), aliquoted, and stored at −20 °C until further use.

RNAi experiments were performed on a randomized pool of nymphs reared from three engorged female ticks collected from the wild. Five microlitres of the *dae2* or *gfp* dsRNA (3 μg μl^{-1}) was loaded into capillary tubes, and 0.5 μl dsRNA was microinjected into the gut of each unfed nymph, as recently described[47]. Microinjected ticks were allowed to rest in a temperature-controlled humidity chamber for 16 h and ~100 nymphs were subsequently fed on *B. burgdorferi*-infected female 4–6-week-old C3H/HeN mice. Two mice were included per RNAi treatment to account for potential variability in *B. burgdorferi* infection loads. Ticks were allowed to feed to repletion (3–5 days) and collected within 24 h ($t = 0$). Fed ticks were maintained in a temperature-controlled incubator until the indicated time point (2 weeks). Knockdown efficiency was analysed by qRT–PCR analysis of *dae2* levels

in engorged nymphs. The RNAi treatment groups were blinded from the time of dsRNA injections through qPCR analyses.

qPCR analyses. qPCR was performed on cDNA samples using the SsoAdvanced Universal SYBR Green Supermix (Biorad). Expression levels for *dae* genes were normalized to actin or *gapdh* expression levels in *I. scapularis* and *N. gruberi*, respectively. Analyses of *dae* gene expression include three technical replicates for *N. gruberi* and technical duplicates of three biological replicates for *I. scapularis*. Populations of *B. burgdorferi* and total bacteria were quantified by qPCR in tick DNA samples using primers targeted to *flaB* (*B. burgdorferi*-specific), the 16S rRNA gene[48], or *TROSPA* (tick-specific). Biological replicates are shown for qPCR analyses of *B. burgdorferi* and total bacterial levels. Transcript or DNA copy numbers were calculated using a standard curve.

28. Swart, E. C. *et al.* The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol.* **11,** e1001473 (2013).
29. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
30. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
31. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
32. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21,** 2104–2105 (2005).
33. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
34. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
35. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
36. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8,** 785–786 (2011).
37. Fulton, C. Amebo-flagellates as research partners: the laboratory biology of Naegleria and Tetramitus. *Methods Cell Biol.* **4,** 341–476 (1970).
38. Fulton, C. & Dingle, A. D. Appearance of the flagellate phenotype in populations of Naegleria amebae. *Dev. Biol.* **15,** 165–191 (1967).
39. Edwards, D. H. & Donachie, D. W. in *Bacterial Growth and Lysis: Metabolism and Structure of the Bacterial Sacculus* (eds de Pedro, M. A., Holtje, J. V. & Loffelhardt, W.) (Plenum Press, 1993).
40. Glauner, B. Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal. Biochem.* **172,** 451–464 (1988).
41. Casjens, S. *et al.* A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi. Mol. Microbiol.* **35,** 490–516 (2000).
42. Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi. Nature* **390,** 580–586 (1997).
43. Jutras, B. L., Chenail, A. M. & Stevenson, B. Changes in bacterial growth rate govern expression of the *Borrelia burgdorferi* OspC and Erp infection-associated surface proteins. *J. Bacteriol.* **195,** 757–764 (2013).
44. Hood, R. D. *et al.* A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7,** 25–37 (2010).
45. Chou, S. *et al.* Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the type VI secretion system. *Cell Reports* **1,** 656–664 (2012).
46. Ramamoorthi, N. *et al.* The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature* **436,** 573–577 (2005).
47. Pal, U. *et al.* TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi. Cell* **119,** 457–468 (2004).
48. Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148,** 257–266 (2002).
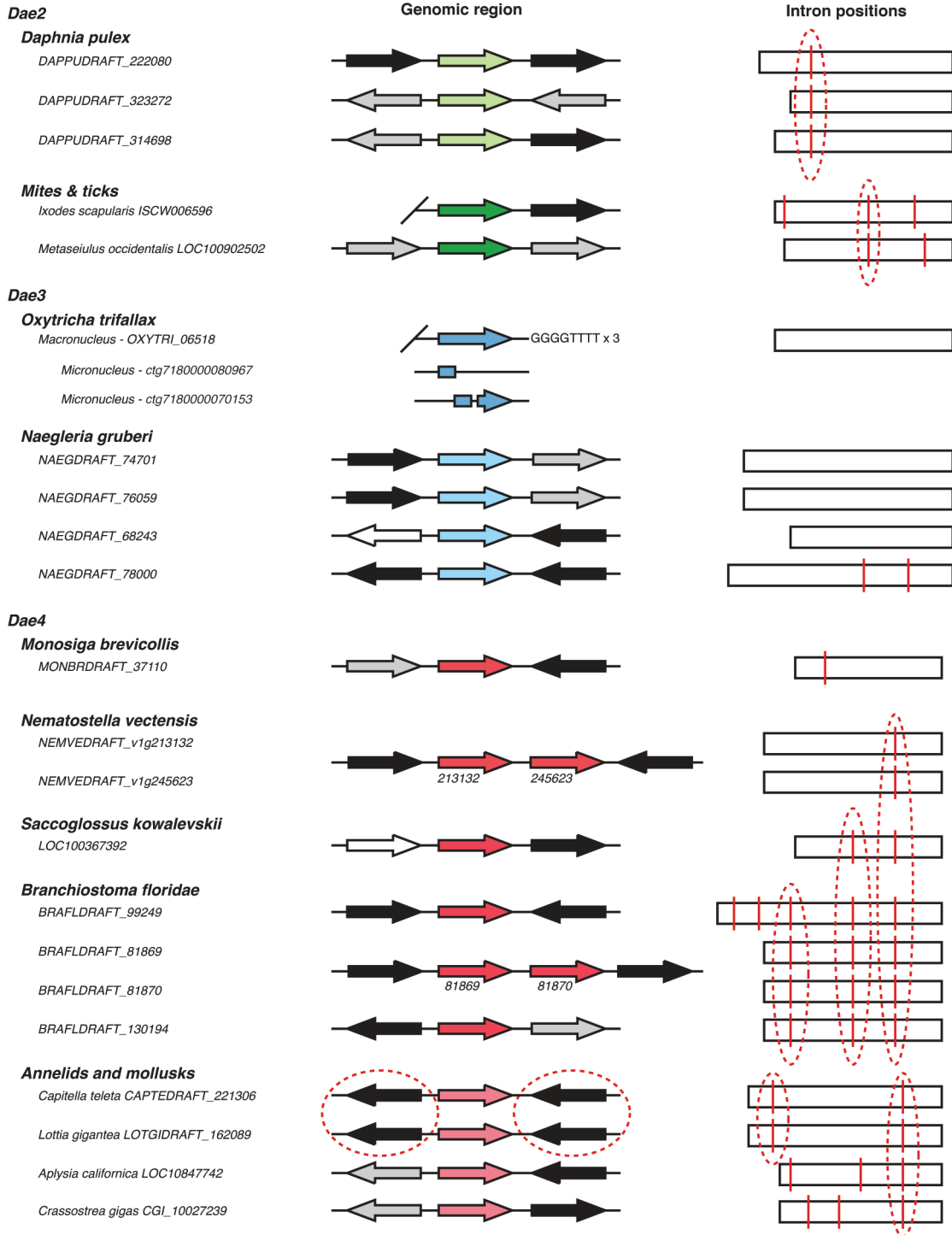
Extended Data Figure 1 | Genomic evidence for validated eukaryotic *dae* genes. Eukaryotic *dae* genes from the indicated organisms are listed adjacent to schematic representations of available predicted open reading frames (colour-coded according to family as in Fig. 1a) and corresponding genomic context of *dae* genes. Flanking genes are colour-coded according to organisms that homologues of these genes are found in (broadly, in eukaryotes, black; only closely related eukaryotic species, grey; both bacteria and eukaryotes, white). Diagonal lines denote ends of genomic contigs. In the right column, splice sites (red vertical lines) and conserved intron positions (red dashed circles) are shown. In *Oxytricha trifallax*, the somatic nucleus (macronucleus) contains ~16,000 chromosomes and is a rearranged form of the germline nucleus (micronucleus)[28]. The complete *dae3* gene in *Oxytricha* is found in the macronucleus on a chromosome with three characteristic GGGGTTTT telomere sequences. Three fragments comprising the *dae3* gene are found in the micronuclear genome (http://oxy.ciliate.org/). In *Nematostella vectensis* and *Branchiostoma floridae*, lineage-specific duplication events have resulted in two adjacent *dae4* paralogues with gene names labelled (numbers). In *Capitella teleta* and *Lottia gigantea*, shared synteny on both sides of the *dae4* gene is indicated (red dashed circles).
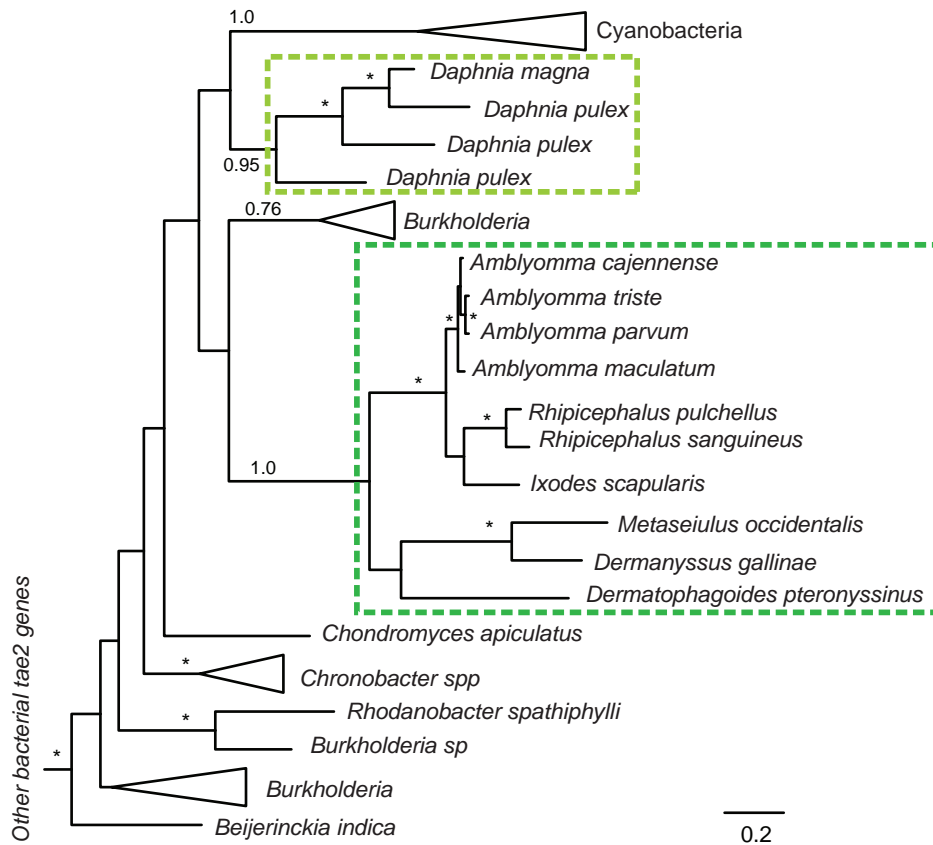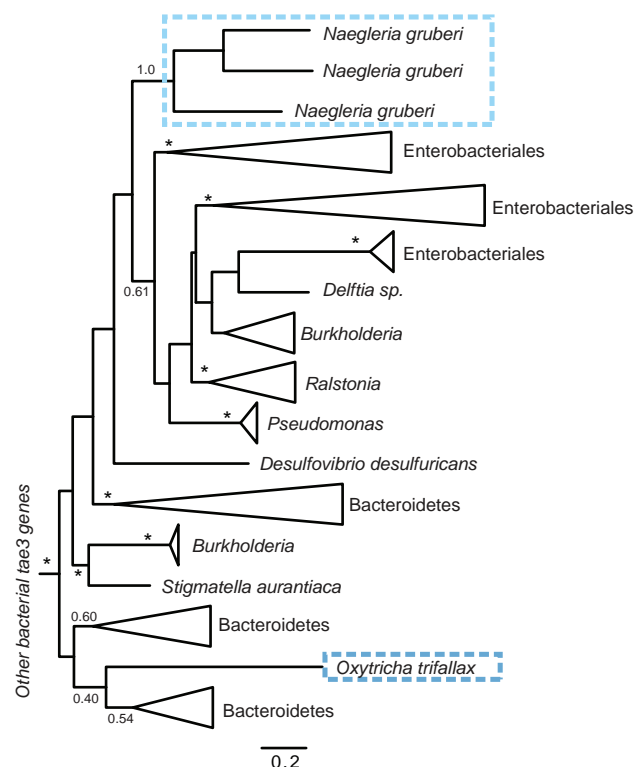
Extended Data Figure 2 | Phylogenetic tree of bacterial *tae2* and eukaroytic *dae2* genes. A phylogenetic tree was constructed using Bayesian methods in MrBayes[34] to compare to the maximum likelihood tree shown in Fig. 1b. Branch support >0.7 is indicated by asterisks or by numbers. The scale bar shows estimated divergence in amino acid changes per residue. Eukaryotic *dae2* genes are indicated by dashed boxes, which highlight two separate HGT events.

In both phylogenetic trees, the two eukaryotic *dae2* clades are well supported as monophyletic clades, supporting our conclusion of two HGT events. Likewise, many major bacterial groups are well supported in both trees. Differences in the overall topology of the trees, mostly owing to changes in deep branches that are not well supported in either phylogenetic tree, reflect uncertainty in the ancient history of these genes and should therefore be treated with caution.
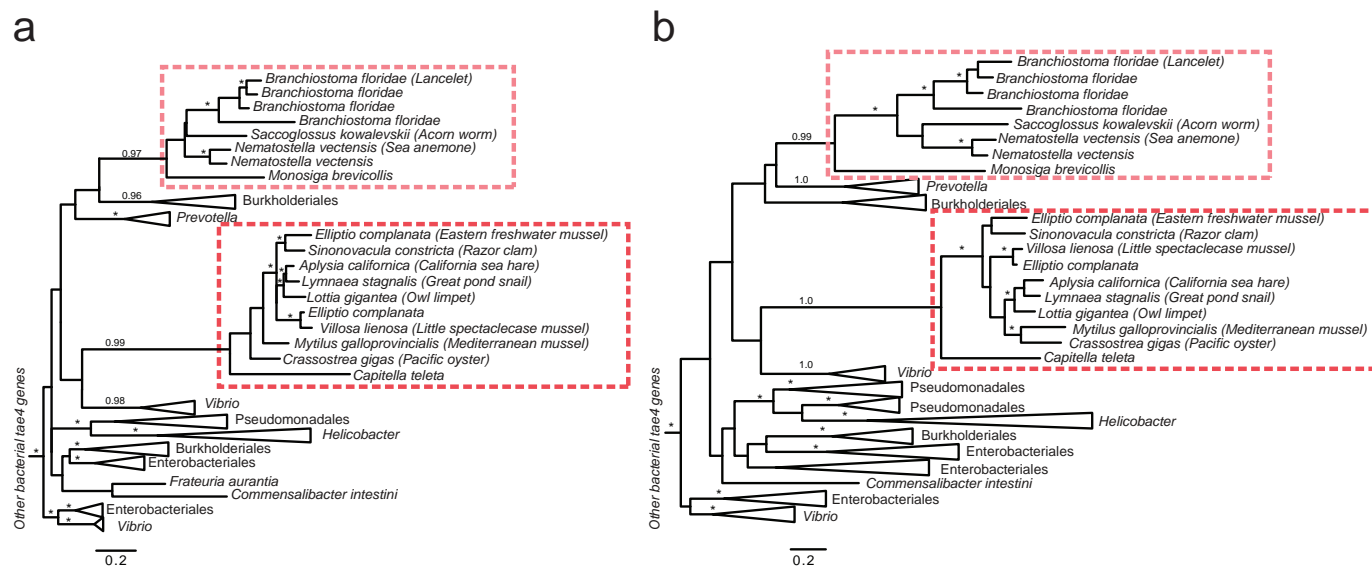
a

b
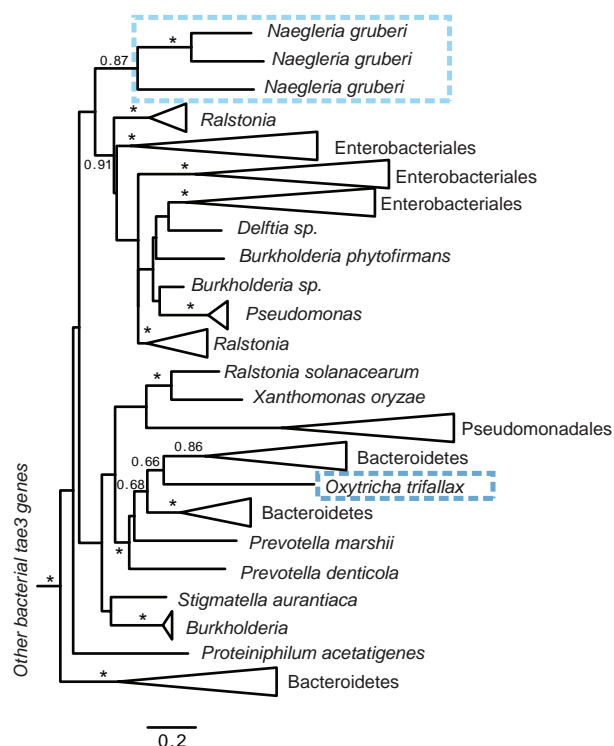
Extended Data Figure 3 | Phylogenetic tree of bacterial *tae3* and eukaryotic *dae3* genes. a, b, Phylogenetic trees were constructed using either maximum likelihood methods (**a**) or Bayesian methods (**b**). Branch support >0.7 is indicated by asterisks or by numbers. The scale bar shows estimated divergence in amino acid changes per residue. Eukaryotic *dae3* genes are indicated by dashed boxes, which highlight two separate HGT events. In both trees, the two

eukaryotic *dae3* clades are well supported as monophyletic clades, supporting our conclusion of two separate HGT events. Likewise, many major bacterial groups are well supported in both trees. Differences in the overall topology of the trees, mostly owing to changes in deep branches that are not well supported in either phylogenetic tree, reflect uncertainty in the ancient history of these genes and should therefore be treated with caution.

Extended Data Figure 4 | Phylogenetic tree of bacterial *tae4* and eukaryotic *dae4* genes. a, b, Phylogenetic trees were constructed using either maximum likelihood methods (**a**) or Bayesian methods (**b**). Branch support >0.7 is indicated by asterisks or by numbers. The scale bar shows estimated divergence in amino acid changes per residue. Eukaryotic *dae4* genes are indicated by dashed boxes, which highlight two separate HGT events. In both trees, the two eukaryotic *dae4* clades are well supported as monophyletic clades, supporting our conclusion of two separate HGT events. Likewise, many major bacterial groups are well supported in both trees. Differences in the overall topology of the trees, mostly owing to changes in deep branches that are not well supported in either phylogenetic tree, reflect uncertainty in the ancient history of these genes and should therefore be treated with caution.
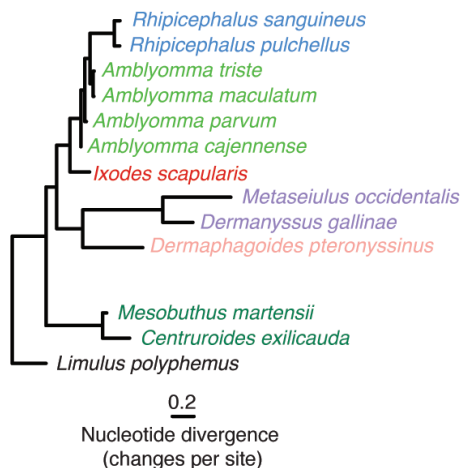
a

```
Ixodes scapularis       ...RWVRGQH|VKSNCGSIPKFTAIATFLKPGNKYLGHAAIFES CASDGIW|VYDQWN...
Metaseiulus occidentalis ...SWRKGAK|VRGICNSIPRSTPVATFVG--KKFRGHAGIFVR|CSRDGIH IYDQTP...
Mesobuthus martensii         \VMSHCSRIPVWTAIATFLREGNLYRGHAAIFES CANDGIW|VYDQTP...
Centruroides exilicausa      \VLSECFKIPIWTAIATFLREDGLYKGHAAIFES CANDGIW|VYDQTP...
Limulus polyphemus           \VKENCKQIPRWTAIATFLSFGDK*GGHSAMLES CDNTGIW|VYDQRP...
                             *      *   **  *.:***:        **:.:: *   ** :***
```

b

Dae2 gene phylogeny

Rhipicephalus sanguineus
Rhipicephalus pulchellus
Amblyomma triste
Amblyomma maculatum
Amblyomma parvum
Amblyomma cajennense
Ixodes scapularis
Metaseiulus occidentalis
Dermanyssus gallinae
Dermaphagoides pteronyssinus
Mesobuthus martensii
Centruroides exilicauda
Limulus polyphemus

0.2
Nucleotide divergence
(changes per site)

c

Chelicerate phylogeny

Metastriate ticks (*Rhipicephalus*)
Metastriate ticks (*Amblyomma*)
Prostriate ticks (*Ixodes*)
Parasitiform mites (*Metaseiulus* and *Dermanyssus*)
Sarcoptiform mites (*Dermatophagoides*)
Trombidiform mites (*Tetranychus*)
Spiders (*Stegodyphus*)
Scorpions (*Mesobuthus* and *Centruroides*)
Horseshoe crab (*Limulus*)

Other arthropods

0   100   200   300   400   500   600
Divergence time
(million years ago)

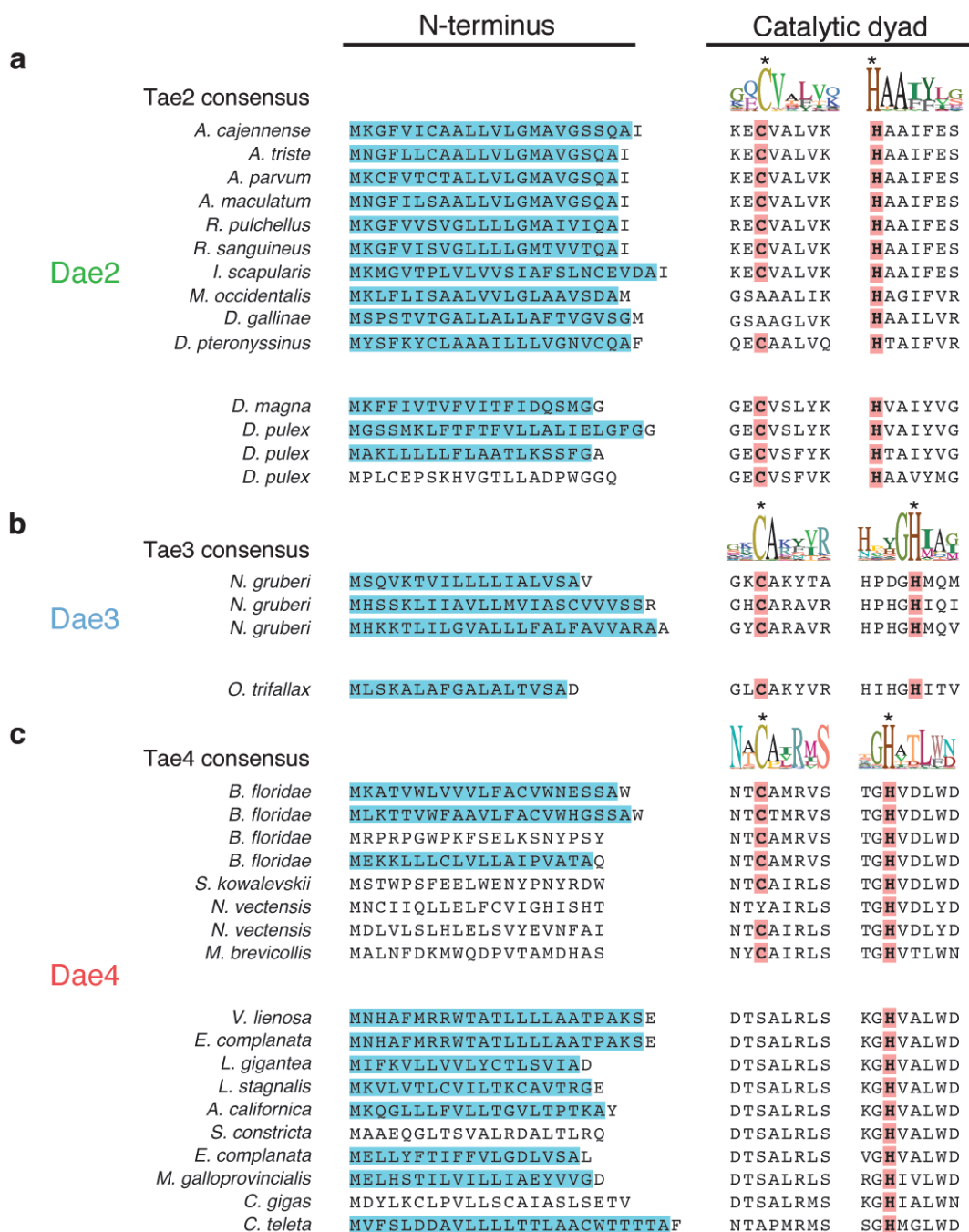**Extended Data Figure 5 | Evidence for *dae2* in other chelicerates.**
a, Alignment of Dae2 from ticks and mites (*I. scapularis* and *M. occidentalis*) with Dae2 sequences from partially assembled genomes of two scorpions (*Mesobuthus martensii* and *Centruroides exilicauda*) and the horsehoe crab (*Limulus polyphemus*). Splice junctions are denoted (vertical red lines). All three alignable partial sequences start (red diagonal slashes) in the same position as the shared splice site in tick and mite *dae2* genes, suggesting that this is probably the beginning of the exons in all *dae* genes shown. A second intron position is shared between the tick, scorpion and horseshoe crab *dae* genes and is nearby the mite intron position. **b**, Phylogenetic tree based on partial nucleotide sequences of *dae2* from the indicated chelicerate species. Scale bar shows estimated divergence, in substitutions per nucleotide. **c**, Chelicerate phylogeny with approximate dates of divergence[13]. The unknown divergence time of sarcoptiform and trombidiform mites is indicated by a question mark. We find no evidence for *dae2* in the complete genome of the trombidiform mite *Tetranychus urticae* nor in the partial (several species) or complete genome (*Stegodyphus mimosarum*) of any spider. Putative *dae2* gene loss events in trombidiform mites and spiders are denoted (dashed lines).

N-terminus   Catalytic dyad

a

Tae2 consensus

Species	N-terminus	Catalytic dyad	
A. cajennense	MKGFVICAALLVLGMAVGSSQAI	KECVALVK	HAAIFES
A. triste	MNGFLLCAALLVLGMAVGSQAI	KECVALVK	HAAIFES
A. parvum	MKCFVTCTALLVLGMAVGSQAI	KECVALVK	HAAIFES
A. maculatum	MNGFILSAALLVLGMAVGSQAI	KECVALVK	HAAIFES
R. pulchellus	MKGFVVSVGLLLLGMAIVIQAI	RECVALVK	HAAIFES
R. sanguineus	MKGFVISVGLLLLGMTVVTQAI	KECVALVK	HAAIFES
I. scapularis	MKMGVTPLVLVVSIAFSLNCEVDAI	KECVALVK	HAAIFES
M. occidentalis	MKLFLISAALVVLGLAAVSDAM	GSAAALIK	HAGIFVR
D. gallinae	MSPSTVTGALLALLAFTVGVSGM	GSAAGLVK	HAAILVR
D. pteronyssinus	MYSFKYCLAAAILLLVGNVCQAF	QECAALVQ	HTAIFVR

Dae2

D. magna	MKFFIVTVFVITFIDQSMGG	GECVSLYK	HVAIYVG
D. pulex	MGSSMKLFTFTFVLLALIELGFGG	GECVSLYK	HVAIYVG
D. pulex	MAKLLLLLFLAATLKSSFGA	GECVSFYK	HTAIYVG
D. pulex	MPLCEPSKHVGTLLADPWGGQ	GECVSFVK	HAAVYMG

b

Tae3 consensus

Species	N-terminus	Catalytic dyad	
N. gruberi	MSQVKTVILLLLIALVSAV	GKCAKYTA	HPDGHMQM
N. gruberi	MHSSKLIIAVLLMVIASCVVVSSR	GHCARAVR	HPHGHIQI
N. gruberi	MHKKTLILGVALLLFALFAVVARAA	GYCARAVR	HPHGHMQV

Dae3

O. trifallax	MLSKALAFGALALTVSAD	GLCAKYVR	HIHGHITV

c

Tae4 consensus

Species	N-terminus	Catalytic dyad	
B. floridae	MKATVWLVVVLFACVWNESSAW	NTCAMRVS	TGHVDLWD
B. floridae	MLKTTVWFAAVLFACVWHGSSAW	NTCTMRVS	TGHVDLWD
B. floridae	MRPRPGWPKFSELKSNYPSY	NTCAMRVS	TGHVDLWD
B. floridae	MEKKLLLCLVLLAIPVATAQ	NTCAMRVS	TGHVDLWD
S. kowalevskii	MSTWPSFEELWENYPNYRDW	NTCAIRLS	TGHVDLWD
N. vectensis	MNCIIQLLELFCVIGHISHT	NTYAIRLS	TGHVDLYD
N. vectensis	MDLVLSLHLELSVYEVNFAI	NTCAIRLS	TGHVDLYD
M. brevicollis	MALNFDKMWQDPVTAMDHAS	NYCAIRLS	TGHVTLWN

Dae4

V. lienosa	MNHAFMRRWTATLLLLAATPAKSE	DTSALRLS	KGHVALWD
E. complanata	MNHAFMRRWTATLLLLAATPAKSE	DTSALRLS	KGHVALWD
L. gigantea	MIFKVLLVVLYCTLSVIAD	DTSALRLS	KGHVALWD
L. stagnalis	MKVLVTLCVILTKCAVTRGE	DTSALRLS	KGHVALWD
A. californica	MKQGLLLFVLLTGVLTPTKAY	DTSALRLS	KGHVALWD
S. constricta	MAAEQGLTSVALRDALTLRQ	DTSALRLS	KGHVALWD
E. complanata	MELLYFTIFFVLGDLVSAL	DTSALRLS	VGHVALWD
M. galloprovincialis	MELHSTILVILLIAEYVVGD	DTSALRLS	RGHIVLWD
C. gigas	MDYLKCLPVLLSCAIASLSETV	DTSALRMS	KGHIALWN
C. teleta	MVFSLDDAVLLLLTTLAACWTTTTAF	NTAPMRMS	SGHMGLWD

**Extended Data Figure 6 | Evidence for retention of important catalytic motifs and recurrent eukaroytic-specific addition of secretion signals.** a–c, Alignments for the predicted Dae N-terminal signal sequences (shaded blue) and catalytic motifs (right) are shown for each of the families. The consensus sequence logo of residues surrounding the cysteine and histidine positions of catalytic dyads from extant Tae enzymes are shown above alignments from each family. Below are aligned eukaroytic Dae proteins in these same regions. Representatives derived from distinct HGT events are separated by a space. Predicted N-terminal secretion signals (blue) and predicted catalytic residues (red) are coloured. Lowering the cut-off value in SignalP[36] from the default value of 0.45 to the 'sensitive' value of 0.34 predicted a signal peptide in residues 1–21 of *C. gigas* Dae4.

a

B. subtilis

b
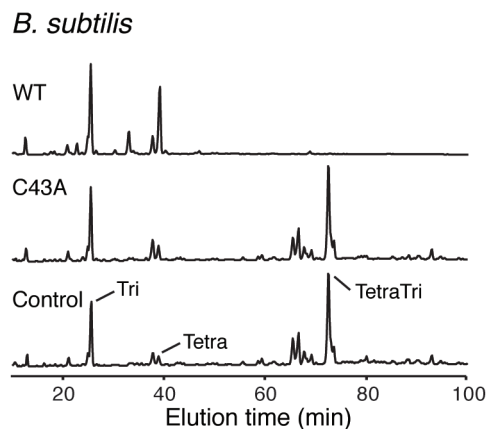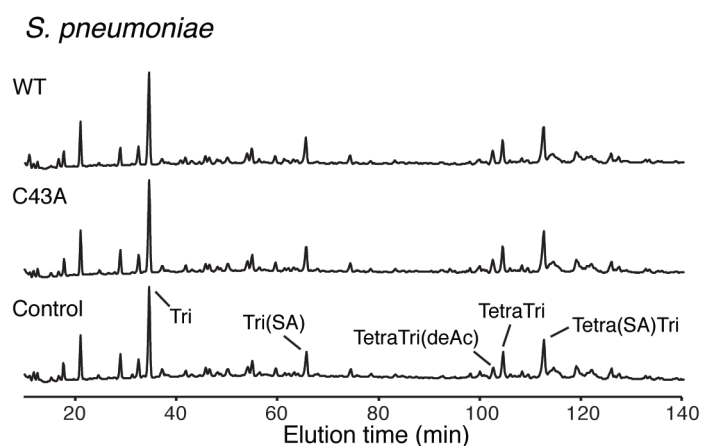
S. pneumoniae

Extended Data Figure 7 | Dae2 degrades *m*DAP- but not Lys-type PG.
a, **b**, Partial HPLC chromatograms of sodium-borohydride-reduced soluble PG fragments (muropeptides) from *Bacillus subtilis* (**a**) or *Streptococcus pneumoniae* (**b**). PG sacculi products resulting from incubation with buffer (Control) or the indicated Dae2 proteins (wild type (WT) or C43A), followed by cellosyl digestion are shown. Major peaks are labelled. **a**, Muropeptides from *B. subtilis* include Tri (GlcNAc–MurNAc(reduced (r))–L-Ala–D-γ-Glu–*m*DAP(amidated (NH$_2$))), Tetra (GlcNAc–MurNAc(r)–L-Ala–D-γ-Glu–*m*DAP(NH$_2$)–D-Ala), and TetraTri (GlcNAc–MurNAc–L-Ala–D-γ-Glu–*m*DAP(NH$_2$)–D-Ala–*m*DAP(NH$_2$)–D-γ-Glu–L-Ala–MurNAc(r)–GlcNAc). **b**, Muropeptides from *S. pneumoniae* include Tri (GlcNAc–MurNAc(r)–L-Ala–D-γ-Gln–L-Lys) and TetraTri (GlcNAc–MurNAc–L-Ala–D-γ-Gln–L-Lys–D-Ala–L-Lys–D-γ-Gln–L-Ala–MurNAc(r)–GlcNAc). L-Ser–L-Ala branch is indicated by '(SA)' and deacetylation by '(deAc)'.
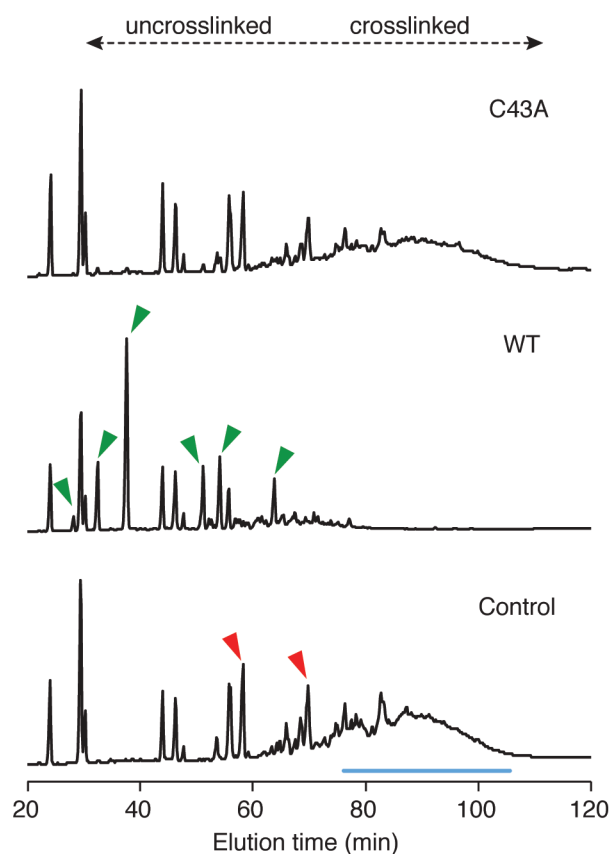
Extended Data Figure 8 | Dae2 is active against *B. burgdorferi* PG. HPLC elution profiles of *B. burgdorferi* sacculi incubated with buffer (Control) or the indicated Dae2 proteins (wild type (WT) or C43A), followed by cellosyl digestion are shown. Discrete peaks lost (red) or produced (green) upon digestion by Dae2 are denoted with arrowheads in control and wild-type chromatograms, respectively. Unresolved peaks, probably corresponding to a complex mixture of multi-crosslinked species cleaved by Dae2, are also highlighted (blue line). *B. burgdorferi* PG composition is complex and not yet resolved, thus approximate elution times of uncrosslinked versus crosslinked species are based on *E. coli* muropeptides in the same solvent system.

a

b

c
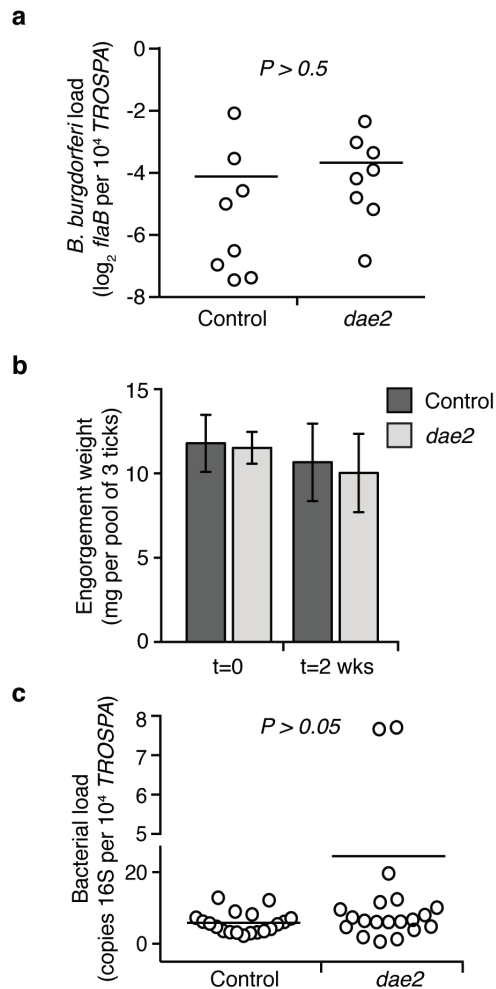
Extended Data Figure 9 | Disruption of *dae2* expression does not significantly alter tick physiology at repletion. a, Knockdown of *dae2* does not increase the *B. burgdorferi* burden in infected nymphs at engorgement. Loads were quantified by qPCR analysis of *flaB*, a *B. burgdorferi*-specific gene, and normalized to *TROSPA*, a tick-specific gene. *n* = 20. For this and subsequent panels, each data point represents a pool of three nymphs, and horizontal bars represent mean values, which were not significantly different in a two-tailed nonparametric Mann–Whitney test ($P > 0.5$). **b**, Disruption of *dae2* expression did not affect engorgement weights of nymphal ticks fed on *B. burgdorferi*-infected mice. Tick weights were measured at repletion and 2 weeks post-repletion. Error bars show ± s.d., *n* = 8. **c**, Overall bacterial load was not affected by knockdown of *dae2*. Bacterial load was assessed by qPCR analysis of the 16S rRNA normalized against the tick-specific gene *TROSPA*. Load is represented on both a linear (bottom) and $\log_2$ (top) scale, which is denoted by a gap on the *y*-axis.

Extended Data Table 1 | **Evolutionary analyses of *dae* and *tae* gene families**

	Gene family	Species group	Number of species	dN/dS	Codons evolving under purifying selection
Eukaryotic amidases	*dae2*	Ticks & mites	10	0.20	21% (25 of 120)
	dae4	Mollusks	9	0.18	40% (59 of 149)
Prokaryotic amidases	*tae2*	*Cronobacter*	10	0.12	23% (30 of 129)
	tae3	*Acinetobacter*	10	0.08	30% (45 of 150)
	tae4	*Pseudomonas*	12	0.15	46% (75 of 163)

Summary of results from maximum likelihood tests of aligned *dae* or *tae* sequences from the indicated species, using SLAC in the HyPhy software package[35]. The overall gene dN/dS ratio (ratio of non-synonymous changes to synonymous changes) is shown, indicating an overall signature of purifying selection. Individual codons with a statistically significant signature of purifying selection ($P < 0.05$) were also calculated and are expressed as a percentage of the total number of codons used in the analysis. In the same analyses, no codons were found with a statistically significant signature of positive selection.

LETTER

Exome sequencing identifies rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction

A list of authors and their affiliations appears at the end of the paper

Myocardial infarction (MI), a leading cause of death around the world, displays a complex pattern of inheritance[1,2]. When MI occurs early in life, genetic inheritance is a major component to risk[1]. Previously, rare mutations in low-density lipoprotein (LDL) genes have been shown to contribute to MI risk in individual families[3–8], whereas common variants at more than 45 loci have been associated with MI risk in the population[9–15]. Here we evaluate how rare mutations contribute to early-onset MI risk in the population. We sequenced the protein-coding regions of 9,793 genomes from patients with MI at an early age (≤50 years in males and ≤60 years in females) along with MI-free controls. We identified two genes in which rare coding-sequence mutations were more frequent in MI cases versus controls at exome-wide significance. At low-density lipoprotein receptor (*LDLR*), carriers of rare non-synonymous mutations were at 4.2-fold increased risk for MI; carriers of null alleles at *LDLR* were at even higher risk (13-fold difference). Approximately 2% of early MI cases harbour a rare, damaging mutation in *LDLR*; this estimate is similar to one made more than 40 years ago using an analysis of total cholesterol[16]. Among controls, about 1 in 217 carried an *LDLR* coding-sequence mutation and had plasma LDL cholesterol > 190 mg dl^{-1}. At apolipoprotein A-V (*APOA5*), carriers of rare non-synonymous mutations were at 2.2-fold increased risk for MI. When compared with non-carriers, *LDLR* mutation carriers had higher plasma LDL cholesterol, whereas *APOA5* mutation carriers had higher plasma triglycerides. Recent evidence has connected MI risk with coding-sequence mutations at two genes functionally related to *APOA5*, namely lipoprotein lipase[15,17] and apolipoprotein C-III (refs 18, 19). Combined, these observations suggest that, as well as LDL cholesterol, disordered metabolism of triglyceride-rich lipoproteins contributes to MI risk.

The US National Heart, Lung, and Blood Institute's exome sequencing project (ESP) sought to use exome sequencing as a tool to identify genes and mechanisms contributing to heart, lung and blood disorders. Within this program, we designed a discovery study for the extreme phenotype of early-onset MI (Fig. 1), as heritability is substantially greater when MI occurs early in life[1,2]. From eleven studies, we identified 1,088 cases with MI at an early age (MI in males ≤50 years old and in females ≤60 years old). As a comparison group, we selected 978 participants from prospective cohort studies who were of advanced age (males ≥60 years old or females ≥70 years old) and free of MI.

We sequenced cases and controls to high coverage by performing solution-based hybrid selection of exons followed by massively parallel sequencing (see Methods)[20]. We performed several quality control steps to identify and remove outlier samples and variants (see Methods and Supplementary Figs 1–13). Characteristics of the discovery set of 1,027 cases and 946 controls are provided in Supplementary Tables 1–3. Across the autosomes, each participant had an average of 43 nonsense, 7,828 missense, 92 splice-site, 189 insertion or deletion (indel) frameshift, 366 indel non-frameshift, and 103 non-synonymous singleton variants.

We first tested whether low-frequency coding variants (defined here as a single nucleotide variant (SNV) or indel with minor allele frequency (MAF) between 1% and 5%) are associated with risk for MI in the discovery sequencing study. We observed no significant association of MI status with any individual variant (Supplementary Fig. 14). We next

evaluated the hypothesis that rare alleles (defined here as a SNV or indel with MAF <1%) collectively within a gene contribute to risk for MI (see Methods). We tested for an excess (or deficit) in cases versus controls of rare, non-synonymous mutations by aggregating together SNVs and indels with MAF <1% ('T1' test) in each gene and comparing the counts in cases and controls[21]. Empirical *P* values were obtained using permutation.

The need to aggregate rare variants requires consideration of which variants to be studied together. Ideally, one would aggregate only harmful alleles and ignore benign alleles. To enrich for harmful alleles, we considered three sets of variants: (1) non-synonymous only; (2) a 'deleterious (PolyPhen)' set consisting of non-synonymous after excluding missense alleles annotated as benign by PolyPhen-2 HumDiv software; and (3) 'disruptive' mutations only (nonsense, indel frameshift, splice-site; also referred to as 'null' mutations). To account for multiple testing, we set exome-wide significance for this study at $P = 8 \times 10^{-7}$, a Bonferroni correction for the testing of ~20,000 genes and three variant sets. When the T1 test was applied across these three sets of alleles in the discovery sequencing study, no gene-based association signal deviated from what we expected by chance (Supplementary Figs 15–22).
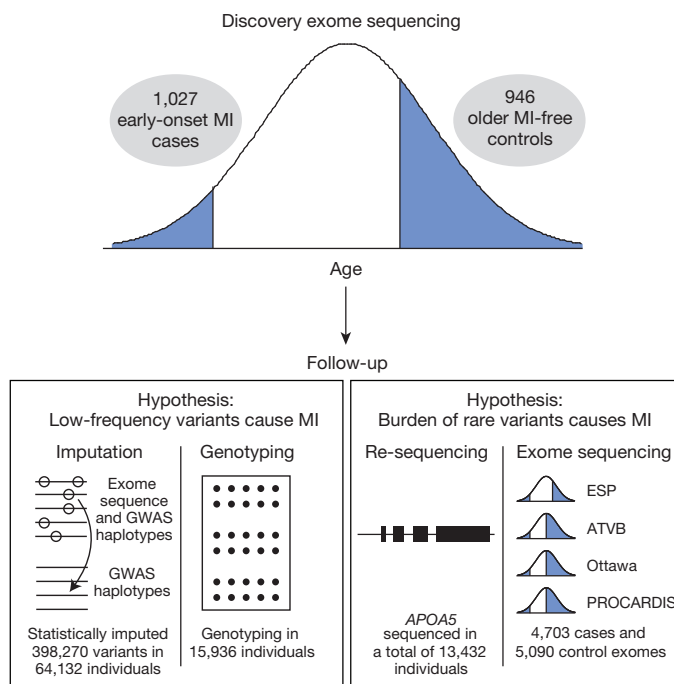

Figure 1 | Overall design for the early-onset myocardial infarction study within the US National Heart, Lung, and Blood Institute's exome sequencing project (ESP). Whole exome sequencing was performed in 1,973 individuals from the phenotypic extremes. To test the hypothesis that low-frequency variants confer risk for myocardial infarction (MI), we performed follow-up statistical imputation and array-based genotyping of single nucleotide variants. To test the hypothesis that a burden of rare mutations in a gene confers risk for MI, we performed targeted re-sequencing and additional exome sequencing.

Table 1 | Association of a burden of rare mutations in *APOA5* with risk for early-onset myocardial infarction or coronary artery disease

Mutation set	*n* cases/controls	T1 cases	T1 controls	Freq cases (%)	Freq control (%)	OR	*P*
Non-synonymous	6,721/6,711	93	42	1.4	0.63	2.2	5×10^{-7}
Deleterious (PolyPhen)	6,721/6,711	63	31	0.94	0.46	2.0	6×10^{-5}
Deleterious (broad)	6,721/6,711	68	31	1.0	0.46	2.2	2×10^{-5}
Deleterious (strict)	6,721/6,711	10	3	0.15	0.045	3.3	0.008
Disruptive	6,721/6,711	9	2	0.13	0.03	4.5	0.007

Summary allele counts and carrier frequencies are shown. Only SNVs and indels with minor allele frequency less than 1% were considered in burden analysis. Deleterious (PolyPhen) as defined by nonsense, splice-site, indel frameshift, and missense annotated as 'possibly damaging' or 'probably damaging' by PolyPhen-2 HumDiv software; 'deleterious (broad)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by at least one of the five protein prediction algorithms of LRT score, MutationTaster, PolyPhen-2 HumDiv, PolyPhen-2 HumVar and SIFT; 'deleterious (strict)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by all five protein prediction algorithms; Disruptive defined as nonsense, splice-site or indel frameshift; T1: alleles from SNVs or indels with minor allele frequency less than 1%; Freq (%): percentage of cases or controls carrying a T1 allele; OR: odds ratio.

We followed up on discovery sequencing results in four ways: (1) statistical imputation; (2) array-based genotyping using the Illumina HumanExome Beadchip ('Exome' chip); (3) targeted re-sequencing; and (4) additional exome sequencing (Fig. 1). Imputation and array-based genotyping were used to mainly evaluate low-frequency variants, whereas targeted re-sequencing and exome sequencing were used to test the role of rare mutations.

With the first and second follow-up approaches: imputation (*n* = 64,132) and array-based genotyping (*n* = 15,936), respectively, we did not identify novel low-frequency variants associated with MI or coronary artery disease (CAD) (see Methods, Supplementary Tables 4–7 and Supplementary Figs 23–27). The top association results for SNVs from array-based genotyping are shown in Supplementary Table 8.

In the third follow-up approach, we re-sequenced several genes in additional cases and controls (see Methods, Supplementary Table 9). After sequencing the exons of *APOA5* in 6,721 cases and 6,711 controls, we identified 46 unique non-synonymous or splice-site SNVs or indel frameshifts with allele frequency < 1% (Supplementary Table 10). Based on these variants, we observed 93 alleles in cases and 42 alleles in controls (*P* = 5×10^{-7}; Table 1, Fig. 2 and Supplementary Table 10). This burden of rare mutation signal was primarily driven by mutations seen in one or two study participants (Fig. 2 and Supplementary Table 10). Carriers of a rare *APOA5* mutation had a 2.2-fold higher risk for MI/CAD than non-carriers (Table 1).

According to a recent report, consideration of variant sets based on multiple protein prediction algorithms might yield stronger association signals[22]. Therefore, we investigated two additional variant sets: (1) 'deleterious (broad)' as defined by nonsense, splice-site, indel frameshift,
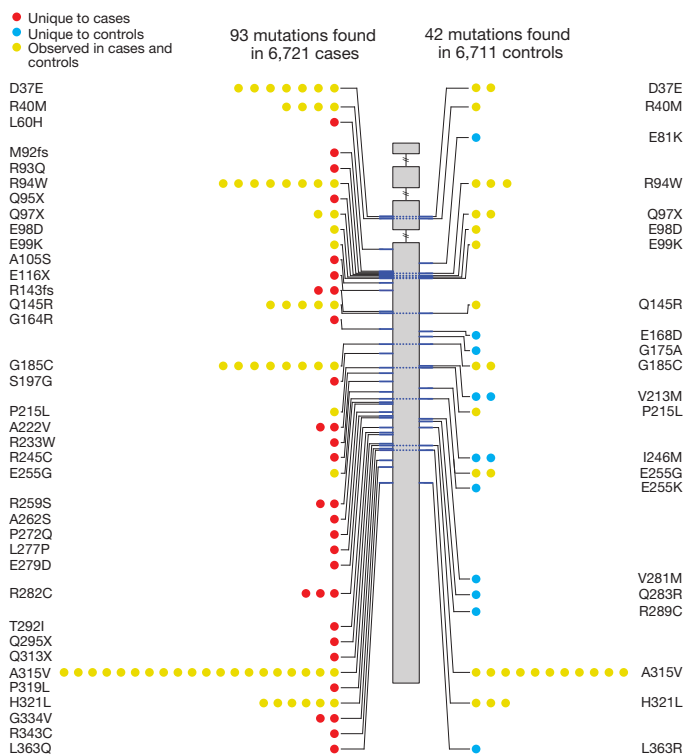

Figure 2 | Apolipoprotein A-V (*APOA5*) mutations discovered after sequencing of 13,432 individuals. Individual mutations (non-synonymous, indel frameshift and splice-site variants with minor allele frequency less than 1%) are depicted according to the genomic position along the length of the *APOA5* gene starting at the 5′ end (top). The number of circles on the left and right represents the number of times that mutation is observed in cases or controls, respectively. Dashed lines across the gene connect the same mutation seen in both cases and controls. Mutations are shaded in red (observed in cases only), blue (observed in controls only) or yellow (observed in both cases and controls).
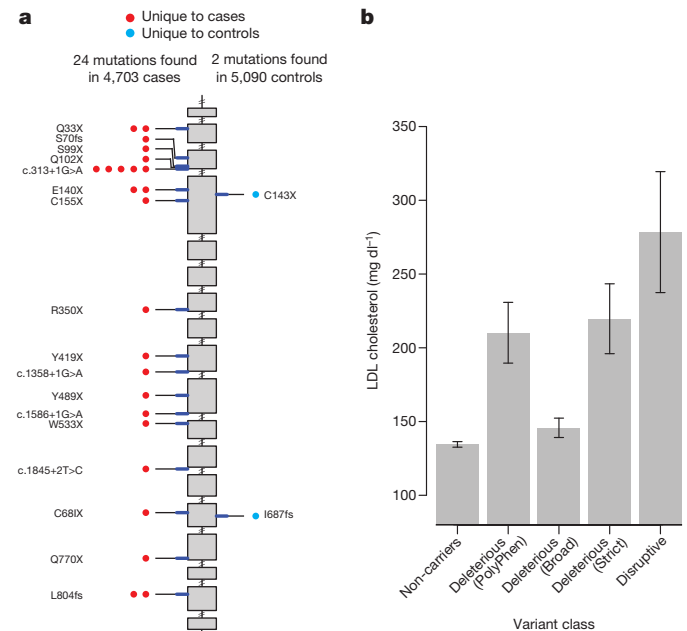

Figure 3 | Low-density lipoprotein receptor (*LDLR*) mutations discovered after sequencing 9,793 individuals. a, Individual disruptive mutations (nonsense, indel frameshift, and splice-site variants with minor allele frequency less than 1%) are depicted according to the genomic position along the length of the *LDLR* gene starting at the 5′ end (top). The number of circles on the left and right represents the number of times that mutation is observed in cases or controls, respectively. Mutations are shaded in red if observed in cases only or blue if observed in controls only. **b,** LDL cholesterol level as observed in different *LDLR* gene mutation annotation categories. Mean (height of bar) and 95% confidence intervals (error bars) are shown. Each individual is categorized based on mutation annotation as follows. Non-carriers: carriers without a missense or disruptive mutation; deleterious (PolyPhen) as defined by nonsense, splice-site, indel frameshift, and missense annotated as 'possibly damaging' or 'probably damaging' by PolyPhen-2 HumDiv software; 'deleterious (broad)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by at least one of five protein prediction algorithms (LRT score, MutationTaster, PolyPhen-2 HumDiv, PolyPhen-2 HumVar and SIFT); 'deleterious (strict)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by all five of the above protein prediction algorithms; disruptive: carriers of mutations that are nonsense, indel frameshift, or splice-site.

Table 2 | Association of a burden of rare mutations in *LDLR* with risk for early-onset myocardial infarction or coronary artery disease

Mutation set	*n* cases/controls	T1 cases	T1 controls	Freq cases (%)	Freq controls (%)	OR	*P*
Non-synonymous	4,703/5,090	285	208	6.1	4.1	1.5	4×10^{-6}
Deleterious (PolyPhen)	4,703/5,090	148	67	3.1	1.3	2.4	1×10^{-11}
Deleterious (broad)	4,703/5,090	243	158	5.2	3.1	1.7	9×10^{-8}
Deleterious (strict)	4,703/5,090	90	23	1.9	0.45	4.2	3×10^{-11}
Disruptive	4,703/5,090	24	2	0.51	0.039	13.0	9×10^{-5}

Summary allele counts and carrier frequencies are shown. Only SNVs and indels with minor allele frequency less than 1% were considered in burden analysis. Deleterious (PolyPhen) as defined by nonsense, splice-site, indel frameshift, and missense annotated as 'possibly damaging' or 'probably damaging' by PolyPhen-2 HumDiv software; 'deleterious (broad)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by at least one of the five protein prediction algorithms of LRT score, MutationTaster, PolyPhen-2 HumDiv, PolyPhen-2 HumVar and SIFT; 'deleterious (strict)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as deleterious by all five protein prediction algorithms; Disruptive defined as nonsense, splice-site or indel frameshift; T1: alleles from SNVs or indels with minor allele frequency less than 1%; Freq (%): percentage of cases or controls carrying a T1 allele; OR: odds ratio.

and missense annotated as damaging by at least one of five protein prediction algorithms; and (2) 'deleterious (strict)' as defined by nonsense, splice-site, indel frameshift, and missense annotated as damaging by all five protein prediction algorithms (see Methods). Carriers of a rare *APOA5* deleterious (strict) mutation had an even higher risk for MI/CAD (3.3-fold, *P* = 0.008).

A burden of rare mutations in *APOA5* explains about 0.14% of the total variance for MI and roughly 0.28% of the heritability (assuming that additive genetic factors explain ~50% of the overall variance) (see Methods and Supplementary Table 11). When compared with non-carriers, carriers of rare non-synonymous *APOA5* alleles had higher plasma triglycerides (median in carriers was 167 mg dl^{-1} versus 104 mg dl^{-1} for non-carriers, *P* = 0.007) and lower high-density lipoprotein cholesterol (mean in carriers was 43 mg dl^{-1} versus 57 mg dl^{-1} for non-carriers, *P* = 0.007), but similar LDL cholesterol (median in carriers was 110 mg dl^{-1} versus 108 mg dl^{-1} for non-carriers, *P* = 0.66) (Supplementary Table 12).

In the fourth follow-up approach, we performed exome sequencing in additional early-onset MI/CAD cases and controls, bringing the total number of exomes analysed to 9,793 (Supplementary Tables 13 and 14). We tested for an excess (or deficit) in cases versus controls of rare mutations in any gene (Supplementary Fig. 28 and Supplementary Tables 15–17). At this sample size, rare alleles collectively conferred risk for MI at exome-wide significance in only one gene, *LDLR* (Fig. 3).

After sequencing the exons of *LDLR* in 4,703 cases and 5,090 controls, we identified 156 unique non-synonymous, splice-site SNVs and indel frameshifts with allele frequency <1% (Table 2 and Supplementary Table 18). Of these variants, we observed 285 alleles in cases (6.1% of cases) and 208 alleles in controls (4.1% of controls) (1.5-fold effect size, *P* = 4×10^{-6}) (Table 2). When restricting analysis to the deleterious (PolyPhen) set, 3.1% of cases and 1.3% of controls carried at least one such rare mutation, for a 2.4-fold effect size (*P* = 1×10^{-11}). A higher effect size of 4.2-fold (*P* = 3×10^{-11}) was observed when restricting to the deleterious (strict) set. When restricting to disruptive alleles, 0.51% of cases and 0.04% of controls carried at least one such rare disruptive mutation, for a 13-fold effect size (*P* = 9×10^{-5}) (Table 2 and Fig. 3).

Among controls, approximately 1 in 217 individuals carried an *LDLR* non-synonymous or disruptive mutation and had LDL cholesterol > 190 mg dl^{-1}; in contrast, among cases, approximately 1 in 51 individuals carried an *LDLR* non-synonymous or disruptive mutation and had LDL cholesterol > 190 mg dl^{-1}.

A burden of rare mutations in *LDLR* explains about 0.24% of the total variance for MI and roughly 0.48% of the heritability (see Methods and Supplementary Table 19). LDL cholesterol level differed based on functional class annotation with the greatest difference seen between carriers of disruptive mutations and those who did not carry any non-synonymous mutations (279 mg dl^{-1} versus 135 mg dl^{-1}, Fig. 3 and Supplementary Table 20). Approximately 49% of the *LDLR* alleles discovered in this study (77 of 156) have been previously observed in *LDLR* familial hypercholesterolemia databases[23] (Supplementary Table 21).

Using these rare variant signals as a guide, we estimated sample sizes that will be required to make similar discoveries. A very large number of samples, at least 10,000 exomes, are required to achieve 80% statistical

power at an exome-wide level of statistical significance (Supplementary Figs 29–31).

Here we show that a burden of rare alleles in two genes, *LDLR* and *APOA5*, contributes to risk for MI. These results suggest several conclusions regarding the inherited basis for MI and rare variant association studies. First, after a DNA sequence-based search across nearly all protein-coding genes in >9,700 early-onset MI cases and controls, *LDLR* is the strongest association signal, with mutations in the gene accounting for about 2% of cases. In 1973, Goldstein and colleagues studied survivors of early MI and noted two common lipid abnormalities: hypercholesterolemia and hypertriglyceridemia[16]. On the basis of a total cholesterol value exceeding ~285 mg dl^{-1}, it was estimated that 4.1% of cases with MI prior to the age of 60 had familial hypercholesterolemia; this original estimate is similar to ours based on direct sequencing. In contrast, the prevalence of harmful *LDLR* mutations in the general population is higher than the original estimate (~0.5 in the present study versus 0.1–0.2% by Goldstein). Second, the rare variant association signal presented here establishes *APOA5* as a bona fide MI gene. Initially discovered through comparative genomics analysis of a region harbouring several lipid regulators (that is, *APOA1* and *APOC3*), the *APOA5* locus harbours common variants associated with plasma triglycerides[24]. Candidate gene and genome-wide association studies have associated common variants at this locus also with MI risk (that is, −1131T>C, *APOA5* promoter region, rs662799, MAF of 8%)[25,26]. However, because of extensive linkage disequilibrium in this region, it had been previously uncertain which gene is responsible for the association with MI. The identification of multiple coding sequence variants within *APOA5* clarifies that this gene contributes to MI risk in the population. Third, these data point to a route to MI beyond LDL cholesterol, namely triglyceride-rich lipoproteins[27] and the lipoprotein lipase pathway. Genetic variation at two other proteins related to *APOA5* function, apolipoprotein C-III (refs 18, 19, 28) and lipoprotein lipase[15,17], has been associated with triglycerides and MI risk. Finally, the present study makes clear that rare variant discovery for complex disease will require the sequencing of thousands of cases and careful statistical analysis. Two reasons for the large sample size requirement are an inability to readily distinguish harmful from benign alleles and the extreme rarity of harmful alleles.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B. & de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* **330,** 1041–1046 (1994).
2. Lloyd-Jones, D. M. *et al.* Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *J. Am. Med. Assoc.* **291,** 2204–2211 (2004).
3. Lehrman, M. A. *et al.* Mutation in LDL receptor: Alu–Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* **227,** 140–146 (1985).
4. Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232,** 34–47 (1986).
5. Soria, L. F. *et al.* Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *Proc. Natl Acad. Sci. USA* **86,** 587–591 (1989).

6. Garcia, C. K. *et al.* Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* **292,** 1394–1398 (2001).
7. Berge, K. E. *et al.* Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* **290,** 1771–1775 (2000).
8. Abifadel, M. *et al.* Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nature Genet.* **34,** 154–156 (2003).
9. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316,** 1488–1491 (2007).
10. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357,** 443–453 (2007).
11. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316,** 1491–1493 (2007).
12. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature Genet.* **41,** 334–341 (2009).
13. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genet.* **43,** 333–338 (2011).
14. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nature Genet.* **43,** 339–344 (2011).
15. The CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genet.* **45,** 25–33 (2013).
16. Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L. & Motulsky, A. G. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* **52,** 1544–1568 (1973).
17. Varbo, A. *et al.* Remnant cholesterol as a causal risk factor for ischemic heart disease. *J. Am. Coll. Cardiol.* **61,** 427–436 (2013).
18. The TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung and Blood Institute *et al.* Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N. Engl. J. Med.* **371,** 22–31 (2014).
19. Jørgensen, A. B., Frikke-Schmidt, R., Nordestgaard, B. G. & Tybjaerg-Hansen, A. Loss-of-function mutations in *APOC3* and risk of ischemic vascular disease. *N. Engl. J. Med.* **371,** 32–41 (2014).
20. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnol.* **27,** 182–189 (2009).
21. Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83,** 311–321 (2008).
22. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506,** 185–190 (2014).
23. Leigh, S. E., Foster, A. H., Whittall, R. A., Hubbart, C. S. & Humphries, S. E. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. *Ann. Hum. Genet.* **72,** 485–498 (2008).
24. Pennacchio, L. A. *et al.* An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294,** 169–173 (2001).
25. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration *et al.* Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet* **375,** 1634–1639 (2010).
26. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).
27. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature Genet.* **45,** 1345–1352 (2013).
28. Pollin, T. I. *et al.* A null mutation in human *APOC3* confers a favorable plasma lipid profile and apparent cardioprotection. *Science* **322,** 1702–1705 (2008).

Supplementary Information is available in the online version of the paper.

Ron Do[1,2,3,4]*, Nathan O. Stitziel[5,6]*, Hong-Hee Won[1,2,3,4]*, Anders Berg Jørgensen[7], Stefano Duga[8], Pier Angelica Merlini[9], Adam Kiezun[4], Martin Farrall[10], Anuj Goel[10], Or Zuk[4], Illaria Guella[8], Rosanna Asselta[8], Leslie A. Lange[11], Gina M. Peloso[1,2,3,4], Paul L. Auer[12], NHLBI Exome Sequencing Project†, Domenico Girelli[13], Nicola Martinelli[13], Deborah N. Farlow[4], Mark A. DePristo[4], Robert Roberts[14], Alexander F. R. Stewart[14], Danish Saleheen[15], John Danesh[15], Stephen E. Epstein[16], Suthesh Sivapalaratnam[17], G. Kees Hovingh[17], John J. Kastelein[17], Nilesh J. Samani[18], Heribert Schunkert[19], Jeanette Erdmann[20], Svati H. Shah[21,22], William E. Kraus[22], Robert Davies[23], Majid Nikpay[23], Christopher T. Johansen[24], Jian Wang[24], Robert A. Hegele[24,25], Eliana Hechter[4], Winfried Marz[26,27,28], Marcus E. Kleber[26], Jie Huang[29], Andrew D. Johnson[30], Mingyao Li[31], Greg L. Burke[32], Myron Gross[33], Yongmei Liu[34], Themistocles L. Assimes[35], Gerardo Heiss[36], Ethan M. Lange[11,37], Aaron R. Folsom[38], Herman A. Taylor[39], Oliviero Olivieri[13], Anders Hamsten[40], Robert Clarke[41], Dermot F. Reilly[42], Wu Yin[42], Manuel A. Rivas[43], Peter Donnelly[43,44], Jacques E. Rossouw[45], Bruce M. Psaty[46,47], David M. Herrington[48], James G. Wilson[49], Stephen S. Rich[50], Michael J. Bamshad[51,52,53], Russell P. Tracy[54], L. Adrienne Cupples[55], Daniel J. Rader[56], Muredach P. Reilly[57], John A. Spertus[58], Sharon Cresci[5,59], Jaana Hartiala[60], W. H. Wilson Tang[61], Stanley L. Hazen[61], Hooman Allayee[60], Alex P. Reiner[12,62], Christopher S. Carlson[12], Charles Kooperberg[12], Rebecca D. Jackson[63], Eric Boerwinkle[64], Eric S. Lander[4], Stephen M. Schwartz[12,62], David S. Siscovick[62,65], Ruth McPherson[23], Anne Tybjaerg-Hansen[7,66], Goncalo R. Abecasis[67], Hugh Watkins[10,43], Deborah A. Nickerson[53], Diego Ardissino[68], Shamil R. Sunyaev[4,69], Christopher J. O'Donnell[29], David Altshuler[1,4], Stacey Gabriel[4] & Sekar Kathiresan[1,2,3,4]

[1]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA [3]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA. [4]Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [5]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. [6]Division of Statistical Genomics, Washington University School of Medicine, St Louis, Missouri 63110, USA. [7]Department of Clinical Biochemistry KB3011, Section for Molecular Genetics, Rigshospitalet, Copenhagen University Hospitals and Faculty of Health Sciences, University of Copenhagen, Copenhagen 1165, Denmark. [8]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milano 20122, Italy. [9]Division of Cardiology, Ospedale Niguarda, Milano 20162, Italy. [10]Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX1 2J, UK. [11]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [12]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [13]University of Verona School of Medicine, Department of Medicine, Verona 37129, Italy. [14]John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario K1Y 4W7, Canada. [15]Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 1TN, UK. [16]MedStar Health Research Institute, Cardiovascular Research Institute, Hyattsville, Maryland 20782, USA. [17]Department of Vascular Medicine, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands. [18]Department of Cardiovascular Sciences, University of Leicester, and Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester LE3 9QP, UK. [19]DZHK (German Research Centre for Cardiovascular Research), Munich Heart Alliance, Deutsches Herzzentrum München, Technische Universität München, Berlin 13347, Germany. [20]Medizinische Klinik II, University of Lübeck, Lübeck 23562, Germany. [21]Center for Human Genetics, Duke University, Durham, North Carolina 27708, USA. [22]Department of Cardiology and Center for Genomic Medicine, Duke University School of Medicine, Durham, North Carolina 27708, USA. [23]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario K1Y 4W7, Canada. [24]Department of Biochemistry, Schulich School of Medicine and Dentistry, Robarts Research Institute, University of Western Ontario, London, Ontario N6A 3K7, Canada. [25]Department of Medicine, Schulich School of Medicine and Dentistry, Robarts Research Institute, University of Western Ontario, London, Ontario N6A 3K7, Canada. [26]Medical Faculty Mannheim, Mannheim Institute of Public Health, Social and Preventive Medicine, Heidelberg University, Ludolf Krehl Strasse 7-11, Mannheim D-68167, Germany. [27]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria.

[28]Synlab Academy, Mannheim 68259, Germany. [29]The National Heart, Lung, Blood Institute's Framingham Heart Study, Framingham, Massachusetts 01702, USA. [30]National Heart, Lung, and Blood Institute Center for Population Studies, The Framingham Heart Study, Framingham, Massachusetts 01702, USA. [31]Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [32]Department of Epidemiology, University of Alabama–Birmingham, Birmingham, Alabama 35233, USA. [33]Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, Minnesota 55455, USA. [34]School of Medicine, Wake Forest University, Winston–Salem, North Carolina 27106, USA. [35]Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. [36]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [37]Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [38]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, Minnesota 55455, USA. [39]University of Mississippi Medical Center, Jackson, Mississippi 39216, USA. [40]Atherosclerosis Research Unit, Department of Medicine, and Center for Molecular Medicine, Karolinska Institutet, Stockholm 171 77, Sweden. [41]Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX1 2JD, UK. [42]Merck Sharp & Dohme Corporation, Rahway, New Jersey 08889, USA. [43]The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX1 2JD, UK. [44]Department of Statistics, University of Oxford, Oxford OX1 2JD, UK. [45]National Heart, Lung, and Blood Institute, Bethesda, Maryland 20824, USA. [46]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, Washington 98195, USA. [47]Group Health Research Institute, Group Health Cooperative, Seattle, Washington 98101, USA. [48]Section on Cardiology, and Public Health Sciences, Wake Forest School of Medicine, Winston–Salem, North Carolina 27106, USA. [49]Jackson Heart Study, University of Mississippi Medical Center, Jackson State University, Jackson, Mississippi 39217, USA. [50]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22904, USA. [51]Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA. [52]Seattle Children's Hospital, Seattle, Washington 98105, USA. [53]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [54]Department of Biochemistry, University of Vermont, Burlington, Vermont 05405, USA. [55]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [56]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [57]Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [58]St Luke's Mid America Heart Institute, University of Missouri-Kansas City, Kansas City, Missouri 64111, USA. [59]Department of Genetics, Washington University in St Louis, Missouri 63130, USA. [60]Department of Preventive Medicine and Institute for Genetic Medicine, University of Southern California Keck School of Medicine, Los Angeles, California 90033, USA. [61]Cardiovascular Medicine, Cleveland Clinic, Cleveland, Ohio 44195, USA. [62]Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA. [63]Ohio State University, Columbus, Ohio 43210, USA. [64]Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA. [65]Department of Medicine, School of Medicine, University of Washington, Seattle, Washington 98195, USA. [66]Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 København N, Denmark. [67]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Missouri 48109, USA. [68]Department of Cardiology, Parma Hospital, Parma 43100, Italy. [69]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.
†A list of authors and their affiliations appears in the Supplementary Information.

METHODS

General overview of the Exome Sequencing Project (ESP). Details of the study design of the National Heart, Lung and Blood Institute's GO exome sequencing project (NHLBI ESP) have been published previously[29]. Briefly, the goal of the NHLBI ESP was to discover rare coding variation in genes contributing to heart, lung and blood disorders using next-generation sequencing of the protein-coding regions of the genome (exome sequencing). The study includes five primary groups including: Seattle GO (University of Washington, Seattle, Washington); Broad GO (Broad Institute, Cambridge, Massachusetts); WHISP (Ohio State University Medical Center, Columbus, Ohio); Lung GO (University of Washington, Seattle, Washington); Heart GO (University of Virginia Health System, Charlottesville, Virginia) and two collaborating groups, WashU GO (Washington University, St Louis) and CHARGE-S GO (University of Texas Health Sciences Center, Houston, Texas).

We included samples from several studies: Women's Health Initiative (WHI); Framingham Heart Study (FHS); Jackson Heart Study (JHS); Multi-Ethnic Study of Atherosclerosis (MESA); Atherosclerosis Risk in Communities (ARIC); Coronary Artery Risk development in Adults (CARDIA); Cardiovascular Health Study (CHS); Lung Health Study (LHS); COPD genetic epidemiology (COPD Gene); severe asthma research project (SARP); pulmonary arterial hypertension (PAH); acute lung injury (ALI); cystic fibrosis (CF); Cleveland Clinic GeneBank (CCGB); Massachusetts General Hospital premature coronary artery disease study (MGH PCAD); Heart Attack Risk in Puget Sound (HARPS); Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status (TRIUMPH) and the PennCath study.

General overview of the ESP early-onset myocardial infarction study. Within the NHLBI ESP, we designed an exome sequencing experiment specifically to study early-onset myocardial infarction (EOMI). We selected EOMI cases and controls from eleven studies, including: ARIC, MESA, CCGB, FHS, HARPS, MGH PCAD, PennCath, TRIUMPH, WHI, CHS, and JHS (Supplementary Tables 1–3). Samples were selected based on the extreme tails of the phenotypic distribution, in order to enrich for a genetic contribution to disease. EOMI cases were defined as individuals who had an MI at an age of ≤50 for men and ≤60 for women. Controls were selected as individuals with no history of MI at baseline or during follow-up to at least age 60 for men and 70 for women. The study samples, along with case and control definitions, are briefly described below and shown in Supplementary Tables 1–3.

Study and phenotype descriptions for ESP EOMI

The HeartGO consortium. HeartGO is a multiethnic consortium consisting of six NHLBI population-based cohorts of men and women: ARIC, CHS, FHS, CARDIA, JHS, and MESA. The age range of participants in these six cohorts spans the spectrum from early adulthood to old age, providing a broad age representation. Each participating cohort in HeartGO has completed ascertainment of multiple phenotypes, including all of the major cardiovascular risk factors (blood pressure, lipids, diabetes status), biomarkers including measures of blood cell counts, subclinical disease imaging, and cardiovascular and lung outcomes including MI and stroke. Participants in all six cohorts provided written informed consent. The NIH database of genotypes and phenotypes (dbGaP) site contains further details regarding the phenotypes accessible for each individual HeartGO cohort.

Cleveland Clinic GeneBank (CCGB). The CCGB study is a single-centre prospective cohort-based study that enrolled patients undergoing elective diagnostic coronary angiography between 2001 and 2006.

Heart Attack Risk in Puget Sound (HARPS). The HARPS study is a population-based case-control study that enrolled cases with incident MI presenting to a network of hospitals in the metropolitan Seattle–Puget Sound region of Washington State between 1998 and 2002.

The Massachusetts General Hospital premature coronary artery disease (MGH PCAD) study. The MGH PCAD study is a hospital-based case-control study that enrolled cases hospitalized with early MI at MGH between 1999 and 2004.

PennCath. The PennCath study is a catheterization-lab based cohort study from the University of Pennsylvania Medical Center and enrolled subjects at the time of cardiac catheterization and coronary angiography between 1998 and 2003. Persons undergoing cardiac catheterization at either the Hospital of the University of Pennsylvania or Penn Presbyterian Medical Center consented for the PennCath study to identify genetic and biochemical factors related to coronary disease.

The Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status (TRIUMPH). The TRIUMPH study is a large, prospective, observational cohort study of consecutive patients with acute MI presenting to 24 US hospitals from April 2005 to December 2008. MI was diagnosed using contemporary definitions[30] and all patients had an elevated troponin blood test.

Women's Health Initiative (WHI). WHI is a major research program that has been ongoing for over 20 years to address the most common causes of death, disability and poor quality of life in postmenopausal women—cardiovascular disease, cancer, and osteoporosis.

Studies involved in follow-up statistical imputation, array-based genotyping, targeted re-sequencing and additional exome sequencing

Statistical imputation. We performed statistical imputation of single nucleotide variants (SNVs) discovered in the exomes of the first 786 samples. We imputed exonic SNVs into 64,132 independent samples in 16 studies to test for association of coding SNVs with MI or CAD. The studies are described in Supplementary Table 5.

Array-based genotyping. We performed follow-up array-based genotyping using the Illumina HumanExome Beadchip ('exome chip') array in 15,936 independent samples from seven studies. The studies are described in Supplementary Table 7.

Targeted re-sequencing. We performed targeted re-sequencing of the *APOA5* gene in an additional 11,414 individuals from five cohorts. The studies are described in Supplementary Table 9.

Exome sequencing-based follow-up. We performed exome sequencing in additional individuals from three cohorts. The studies are described in Supplementary Table 13.

Detailed methods for the processing and analysis of samples for the various stages of the project are described below. We describe methods for the different stages of the project, including discovery exome sequencing, follow-up imputation, array-based genotyping, targeted re-sequencing and additional exome sequencing.

Laboratory methods for discovery exome sequencing in the ESP EOMI Project.

Exome sequencing. Exome sequencing was performed at the Broad Institute. Sequencing and exome capture methods have been previously described[29]. A brief description of the methods is provided below.

Receipt/quality control of sample DNA. Samples were shipped to the Biological Samples Platform laboratory at the Broad Institute of MIT and Harvard. DNA concentration was determined by the Picogreen assay (Invitrogen) before storage in 2D-arcoded 0.75 ml Matrix tubes at −20 °C in the SmaRTStore (RTS, Manchester, UK) automated sample handling system. We performed initial quality control (QC) on all samples involving sample quantification (PicoGreen), confirmation of high-molecular weight DNA and fingerprint genotyping and gender determination (Illumina iSelect). Samples were excluded if the total mass, concentration, integrity of DNA or quality of preliminary genotyping data was too low.

Library construction and in-solution hybrid selection. Starting with 3 μg of genomic DNA, library construction and in-solution hybrid selection were performed as described previously[31]. A subset of samples, however, was prepared using this protocol with some slight modifications. Initial genomic DNA input into shearing was reduced from 3 μg to 100 ng in 50 μl of solution. In addition, for adaptor ligation, Illumina paired-end adapters were replaced with palindromic forked adapters with unique 8 base index sequences embedded within the adaptor.

Preparation of libraries for cluster amplification and sequencing. After in-solution hybrid selection, libraries were quantified using qPCR (KAPA Biosystems) with probes specific to the ends of the adapters. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 2 nM and then denatured using 0.1 N NaOH using Perkin-Elmer's MultiProbe liquid handling platform. A subset of the samples prepared using forked, indexed adapters was quantified using qPCR, normalized to 2 nM using Perkin-Elmer's Mini-Janus liquid handling platform, and pooled by equal volume using the Agilent Bravo. Pools were then denatured using 0.1 N NaOH. Denatured samples were diluted into strip tubes using the Perkin-Elmer MultiProbe.

Cluster amplification and sequencing. Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using either Genome Analyzer v3, Genome Analyzer v4, or HiSeq 2,000 v2 cluster chemistry and flowcells. After cluster amplification, SYBR green dye was added to all flowcell lanes, and a portion of each lane visualized using a light microscope, in order to confirm target cluster density. Flowcells were sequenced either on Genome Analyzer II using v3 and v4 Sequencing-by-Synthesis Kits, then analysed using RTA v1.7.48, or on HiSeq 2,000 using HiSeq 2,000 v2 Sequencing-by-Synthesis Kits, then analysed using RTA v1.10.15. All samples were run on 76 cycle, paired end runs. For samples prepared using forked, indexed adapters, Illumina's Multiplexing Sequencing Primer Kit was also used.

Read mapping and variant analysis. Samples were processed from real-time base-calls (RTA 1.7 software [Bustard], converted to qseq.txt files, and aligned to a human reference (hg19) using Burrows–Wheeler Aligner (BWA, see ref. 32). Aligned reads duplicating the start position of another read were flagged as duplicates and not analysed ('duplicate removal'). Data was processed using the Genome Analysis ToolKit (GATK v1.1.3, ref. 33). Reads were locally realigned (GATK IndelRealigner) and their base qualities were recalibrated (GATK TableRecalibration). Variant detection and genotyping were performed on both exomes and flanking 50 base pairs of intronic sequence using the UnifiedGenotyper (UG) tool from the GATK. Variant data for each sample was formatted (variant call format (VCF)) as 'raw' calls for all samples. SNVs and indel sites were flagged using the variant filtration

walker (GATK) to mark sites of low quality that were likely false positives. SNVs were marked as potential errors if they exhibited strong strand bias (SB $\geq$ 0.10), low average quality (quality per depth of coverage (QD) < 5.0), or fell in a homopolymer run (HRun > 4). Indels were marked as potential errors for low quality (quality score (QUAL) < 30.0), low average quality (QD <2.0), or if the site exhibited strong strand bias (SB > −1.0). Samples were considered complete when exome targeted read coverage was $\geq$ 20× over $\geq$ 80% of the exome target.

Data analysis QC. Fingerprint concordance between sequence data and fingerprint genotypes was evaluated. Variant calls were evaluated on both bulk and per-sample properties: novel and known variant counts, transition–transversion (TS-TV) ratio, heterozygous–homozygous non-reference ratio, and deletion/insertion ratio. Both bulk and sample metrics were compared to historical values for exome sequencing projects at the Broad Institute. No significant deviation of the ESP variants or ESP samples from historical values was noted.

Data processing, quality control and association analysis of discovery exome sequencing

Variant calling. Variants (SNVs and indels) were identified and genotyped from recalibrated BAM files[34] using the multi-sample processing mode of the Unified Genotyper tool from the GATK. Variants were first identified and genotyped in random batches of 100 samples. The batches were then merged into a single VCF file using the GATK CombineVariants tool.

Variant annotation. Variants (SNVs and indels) were annotated using the GRCh37.64 database using the SNP effect predictor tool (SnpEff, see ref. 35) and the GATK VariantAnnotator. The primary SnpEff genomic effects that were annotated include: splice-site acceptor, splice-site donor, indel frameshift, indel non-frameshift, nonsense, non-synonymous and synonymous variants. For variants that have different annotations due to multiple transcripts of the gene, the highest impact effect for each variant was taken.

Sample level quality control. We performed several quality control steps to identify and remove outlier samples (Supplementary Figs 1–8). First, we required that each sample had a minimum of 20-fold coverage for at least 80% of the targeted bases. Second, we compared self-reported ancestry with that inferred from the sequence data and removed discordant samples. Third, we removed samples with high degree of heterozygosity and low number of singleton counts as this pattern suggests DNA contamination across samples. Fourth, we removed samples with an extremely high number of variants or singletons as this can suggest low quality DNA. Finally, we removed samples exhibiting a mismatch between the reported gender and that inferred from sequence data. Of 2,066 cases and controls sequenced across the exome, we removed 93 samples due to these exclusion criteria.

Variant level quality control. QC measures were also performed to remove low quality variants. We assessed population genetics metrics including the TS–TV ratio, the ratio of the number of heterozygous changes to the number of homozygous non-reference changes, and the number of non-synonymous to the number of synonymous changes. This analysis can help filter false positive calls since we expect the true TS–TV to be around ~3.2 in European populations[33], while a set of random SNVs (or false positive variants) should give a random expectation of 0.5. Variants with low depth of coverage (DP) and high percent missingness generally had low TS–TV and heterozygous–homozygous non-reference ratios. Variants were removed if there was DP <8 average per sample and >2% missingness (Supplementary Figs 9–12). Distribution of allele frequencies of the SNVs is shown in Supplementary Fig. 13.

Common variant association analysis. We performed single variant association analysis in our exome sequencing data set. For SNVs with MAF greater than 5%, we ran logistic regression, after adjusting for 10 principal components while for SNVs with MAF less than 5%, we ran Fisher's Exact test. We performed association analysis in European Americans and African Americans separately and then performed sample size weighted meta-analysis using METAL[36]. The association results are shown in Supplementary Fig. 14.

Rare variant association analysis. To test whether rare mutations contribute to MI, we performed burden of rare variant analysis on the ~2,000 ESP EOMI exome samples. We performed a variant of the Combined Multivariate Collapsing test[21], that groups the count of alleles of SNVs in cases and controls. Phenotype labels were permuted 100,000 times to assign a statistical significance. We accounted for ethnicity by permuting phenotype labels within each ethnicity. Association analysis was performed using PLINK/SEQ.

We collapsed variants based on computational predictions from PolyPhen-2 HumDiv[37]. Minor allele frequencies were calculated from all available samples sequenced in each study in order to obtain the most accurate MAF estimates. Therefore, calculation of MAF for ESP EOMI 1 and 2 was performed on a larger set of exome samples that were sequenced at the Broad Institute as part of ESP (n = 970 exomes for ESP EOMI 1 and n = 3,014 for ESP EOMI 2). For our burden of rare variant association analysis, we use a MAF threshold of 1% (T1). Furthermore, we use three different types of variant groupings when collapsing by gene.

These variant groups are: (1) non-synonymous only; (2) a deleterious set consisting of non-synonymous after excluding missense alleles annotated as benign by PolyPhen-2 HumDiv software; and (3) disruptive (nonsense, indel frameshift, splice-site) mutations only. We also performed the T1 test after collapsing all non-synonymous mutations by KEGG pathways (Supplementary Figs 21 and 22).

Methods for follow-up statistical imputation

Construction of reference panels and targeted imputation panels. Exome imputations were performed using two reference panels and 16 targeted imputation panels. A total of 697 ESP samples (436 African Americans and 261 European Americans) were used for the first reference panel while 89 samples from the 1000 Genomes Project[38] were drawn for the second reference panel. For the ESP reference panel, all samples from ARIC (n = 212), JHS (n = 119), MGH PCAD or HARPS (n = 151) and WHI studies (n = 41) were genotyped using commercially available Affymetrix 6.0 arrays. Samples from the FHS (n = 174) were genotyped using the Affymetrix 5.0 array. The second reference panel was comprised of samples from the 1000 Genomes Project that had genotype data for both low coverage sequencing and high coverage exome sequencing data[38]. A total of 89 samples were selected from 6 diverse populations (23 African Ancestry in Southwest US (ASW), 9 Utah residents with Northern and Western European ancestry (CEU), 12 Colombian in Medellin, Colombia (CLM), 25 Mexican Ancestry in Los Angeles, CA (MXL), 17 Toscani in Italia (TSI) and 3 Yoruba in Ibadan, Nigeria (YRI) samples). Low-coverage whole genome sequencing, high-coverage exome sequencing and targeted exome capture were performed based on standard protocols at the Broad Institute. Details of the sequencing methods and samples have been described previously[38]. Imputation was performed into 16 independent study samples with genome-wide genotype data. Study samples were genotyped using commercially available Affymetrix or Illumina genotyping arrays. Further details are described in Supplementary Table 5.

Reference panels were created by merging genotypes from SNVs that span the entire genome (hence, providing a haplotype 'scaffold'), with genotypes from SNVs from ESP exome sequencing data. The first reference panel was generated using genotypes from both genome-wide SNV arrays obtained from dbGAP and exome sequencing data. The second reference panel was generated using genotype data for both low coverage sequencing and high coverage exome sequencing data. Both the reference panel and targeted genome-wide panel were phased using the 'best guess haplotypes' option in IMPUTE2 (ref. 39). Haplotype phasing were performed in 5 megabase chunks as recommended by the software tutorial[39].

Data processing, quality control and association analysis. Imputation of the exome was performed using IMPUTE2. We imputed approximately 400,000 coding SNVs from the reference panels into 28,068 cases and 36,064 controls from 16 different study samples with genome-wide data. Descriptions for the study samples have been reported elsewhere (Supplementary Table 5 for references). We filtered SNVs with MAF < 1% and imputation quality (INFO) <0.5 from further analysis. The distribution of imputation qualities of the SNVs is shown in Supplementary Figs 23 and 24. Association testing for CAD/MI was performed using the score method and assuming an additive model in SNPTEST[40]. Age, sex and the first two principal components were used as covariates when appropriate. We did not observe any indication of excess inflation of test statistics in any of the study samples (Supplementary Table 22). Meta-analysis of study-specific P values for imputed SNVs was performed using the Z-score method weighted by sample size in METAL. Beta and standard errors were estimated based on an inverse-weighted meta-analysis. The distribution of association results for the imputation results is shown in Supplementary Fig. 25 and top association results in Supplementary Table 6.

Methods for follow-up array-based genotyping

Laboratory methods. DNA samples were sent to the Broad Institute Genetic Analysis Platform for genotyping and were placed on 96-well plates for processing using the Illumina HumanExome v1.0 SNP array. Genotypes were assigned using GenomeStudio v2010.3 using the calling algorithm/genotyping module version 1.8.4 along with the custom cluster file StanCtrExChp_CEPH.egt. Only samples passing an overall call rate of 98% criteria and standard identity check were released from the genetic analysis platform.

Data processing, quality control and association analysis. To identify single low-frequency SNVs associated with MI or CAD, we performed array-based genotyping using the Illumina Human Exome Beadchip. We genotyped 83,680 sites identified from exome sequencing in 1,027 early-onset MI cases and 946 controls. The samples for genotyping were drawn from the cohorts listed in Supplementary Table 7 and have been previously described. The functional effect of each variant was predicted using the SeattleSeq Annotation server. For variants having more than one functional class, the most deleterious class was retained.

Several quality control processes were employed to ensure high quality genotypes and samples were used in the association analysis. Samples were excluded for the following criteria: greater than 5% missing genotypes; discordance between inferred gender based on genotype and self-reported gender; inbreeding coefficient less than −0.2 or greater than 0.2; duplicated samples; or proportion of genotypes

identical by descent >0.2. In addition, principal components were calculated using Eigenstrat 4.2 (ref. 41) and samples were removed if they were found to be statistical population outliers. Variants were removed for the following criteria: MAF $= 0\%$; significant difference between missingness in cases compared with controls; extreme deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$); or significant association with genotyping plate assignment. All quality control filtering were performed using PLINK[42] and R (The R Project for Statistical Computing, Vienna, Austria).

Association testing for CAD/MI was performed within each study separately using logistic regression with ten principal components of ancestry as covariates. An inverse standard-error weighted meta-analysis was performed to combine results across studies. The association testing was performed using PLINK[42] and the meta-analysis was performed using METAL. There was no indication of an inflation of test statistics across studies (Supplementary Table 23). The stability of logistic regression was assessed by examining the standard error of the beta estimate as a function of minor allele frequency (see Supplementary Fig. 32). As shown, logistic regression is unstable for a MAF $< 0.05\%$. Fisher's Exact test was used for variants with MAF $< 0.05\%$. The top association results are shown in Supplementary Table 8.

Methods for follow-up re-sequencing
Selection of genes. We first selected six associated genes (based on biologic and/or statistical evidence with T1 $P < 0.005$; *APOA5, CHRM5, SMG7, LYRM1, APOC3, NBEAL1*) for replication sequencing in the ATVB study (Supplementary Table 24) where all cases had suffered an MI before age of 46. We also pursued the same six genes in the Ottawa Heart Study with 552 cases and 586 controls (Supplementary Table 25). One of the genes (*APOA5*) continued to show significant results and was sequenced in three additional studies (**Table 1** and Supplementary Table 26). In total, we performed follow-up sequencing of *APOA5* in six study samples, including the Verona heart study (VHS), Ottawa heart study (OHS), additional exomes from atherosclerosis, thrombosis, and vascular biology Italian study group (ATVB), additional exomes from the ESP EOMI study (ESP EOMI 2), Precocious Coronary Artery Disease Study (PROCARDIS), and the Copenhagen City Heart Study and Copenhagen Ischaemic Heart Disease Study (CCHS/CIHDS).

Laboratory methods. For the VHS study, genomic DNA was extracted from white blood cells using the salting-out method. The protein-coding regions corresponding to the RefSeq transcripts NM_052968 for *APOA5* and NM_012125 for *CHRM5* were sequenced using in-house designed primers (available on request) and the BigDye Terminator Cycle Sequencing Kit v1.1 on an ABI-3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA). SNVs were called using the Variant Reporter software v1.1 (Applied Biosystems).

For the OHS study, PCR primers were designed, tested and optimized to target the exons and flanking non-coding sequences for each gene. Sequencing reactions were performed using big dye terminator chemistry and chromatograms obtained with an Applied Biosystems ABI 3730XL capillary sequencer. Chromatograms were base-called by using Phred, assembled into contigs by using Phrap, and scanned for SNVs with PolyPhred[43] to identify polymorphic sites. Each read was trimmed to remove low-quality sequence (Phred score <25), resulting in analysed reads with an average Phred quality of 40. After assembly and variant calling, each polymorphic site was reviewed by a data analyst using Consed[44] to ensure the quality and accuracy of the variant calls. This process generates sequence-based SNV genotypes with accuracy $>99.9\%$.

For the PROCARDIS study, a single long range PCR product (LRPCR) was amplified to provide coverage of the *APOA5* exonic, intronic and flanking sequences (human reference sequence NCBI build 37 chromosome 11:116,659,905–116,664,331). The LRPCR products were tagged with unique sequence (barcode) adaptors, and processed into 56 short amplicons (Reflex reactions, http://www.populationgenetics.com) and pooled for multiplex next-generation sequencing (NGS). NGS was performed on a MiSeq personal sequencer to $>20\times$ coverage across 95% of the *APOA5* target region on 1,385 MI cases and 1,499 controls. Paired-end reads were mapped to NCBI build 37 using the BWA and SMALT aligners; variants were identified by the GATK unified genotyper (v1.6.13) and annotated using SnpEff v2.0.5 and the GRCh37.64 database.

For the CCHS/CIHDS study, lightscanner screening and re-sequencing were performed. Genomic DNA was isolated from frozen whole blood (QiaAmp4 DNA blood mini kit; QIAGEN, Hilden, Germany). Six PCR fragments were amplified covering the three coding exons and adjacent splice-sites (approximately 20 base pairs upstream and downstream each exon) of *APOA5*. Mutational analysis of the PCR products was performed by high-resolution melting curve (HRM) analysis using the Lightscanner system (Idaho Technology, Salt Lake City, Utah). PCR fragments showing heteroduplex formation by HRM analysis were subsequently sequenced on an ABI 3730 DNA analyser (Applied Biosystems, Foster City, CA).

Data processing, quality control and association analysis. After sequencing, variants were annotated using SnpEff or Annovar[45]. For each study, only non-synonymous SNVs with MAF $<1\%$ were analysed. Rare variant burden testing was performed using the T1 test. Meta-analysis was performed to combine evidence across study specific P values using the sample size weighted Z-score method, implemented in METAL. Association results and a listing of *APOA5* mutations discovered from sequencing are described in **Table 1** and Supplementary Table 10. P values for association between *APOA5* mutation carrier status and lipid traits were performed using the Mann–Whitney ranksum test. Results are shown in Supplementary Table 12.

Methods for follow-up exome sequencing
Laboratory methods. We performed follow-up exome sequencing in additional samples from three other studies. Sequencing was performed at the Broad Institute, using the same protocols described above for the NHLBI ESP Project.

Data processing, quality control and association analysis. Variant calling and annotations were performed as described above for the NHBLI ESP EOMI. Quality control of samples was performed using the following steps. To detect mismatched samples, we calculated discordance rates between genotypes from exome sequencing with genotypes from array-based genotyping. We removed samples with discordance rate > 0.02. We tested for sample contamination using verifyBamID[46], which examines the proportion of non-reference bases at reference sites. We removed samples with FREEMIX or CHIPMIX scores >0.2. Furthermore, we removed outlier samples with too many or too few SNVs (>700 or <5 singletons, >400 or <5 doublets, $>16,000$ ($>20,000$ for African) or $<10,000$ total SNVs), and those with too high or low TS–TV (>4 or <3) and heterozygosity (heterozygote to homozygote non-reference ratio >6 or <2). Finally, we removed samples with high missingness (>0.1). In total, 202 samples were removed. For quality control of variants, we removed SNVs and indels that had low recalibration scores after running GATK VariantRecalibrator. We also removed SNVs with low coverage (DP $<140,000$ and quality over depth (QD) <2) and high missingness (frequency of missing genotypes >0.02). For quality control of indels, we removed indels that had excessive strand bias (Fisher Strand >200), high proportion of alternate alleles seen near the ends of reads (ReadPosRankSum <-20), deviation from Hardy–Weinberg equilibrium (InbreedingCoeff <-0.8) and low coverage (QD <3). Rare variant association analysis was performed using EPACTS. We performed burden of rare variant analysis using the Efficient Mixed-Model Association eXpedited (EMMAX) Combined Multivariate and Collapsing (CMC) test[47]. This approach uses a kinship matrix to take into account population structure. We restricted analyses to SNVs and indels with minor allele frequency <0.01. Furthermore, we restricted analyses to three different sets of variants: (1) non-synonymous only; (2) a deleterious set consisting of non-synonymous after excluding missense alleles annotated as benign by PolyPhen-2 HumDiv software; and (3) disruptive (nonsense, indel frameshift, splice-site) mutations only.

Estimation of heritability explained by a burden of rare mutations in the *APOA5* and *LDLR* genes. We calculated the heritability explained by a burden of rare mutations in the *APOA5* and *LDLR* genes using the following assumptions. We assumed that the alleles come from a mixture of two distributions: harmless alleles, with no effect on the trait, and null alleles, which destroy the function of the gene and have an (constant) effect on the trait. We assumed different values for the fraction of null alleles, α (our current expectation for most genes for α is around one-third to one-half for missense alleles and here we clump missense alleles together with nonsense alleles, which should slightly increase α). The variance explained is sensitive to this parameter. We assumed a liability-threshold model for disease, with an underlying (un-observed) continuous trait representing risk for MI, and MI occurring if risk is above a certain threshold. We assume all null alleles have effect β (in units of standard deviations) on the liability scale. We assumed different values for the prevalence (denoted κ) for early MI (3% to 5%). Results are somewhat sensitive to prevalence; higher prevalence will slightly increase heritability estimates. Given the prevalence, the number of carriers in cases and controls gives us the allele frequency in the population (which is very close to the allele frequency in controls).

We fitted the effect size (β on liability scale) and alleles for different values of α and κ. Results for *APOA5* are shown in Supplementary Table 11 and results for *LDLR* are shown in Supplementary Table 19. For *APOA5*, β is moderate (up to roughly one standard deviation), with variance explained between 0.08% and 0.17% of the total phenotypic variance (on the liability scale). If we assume the heritability of MI is 50%, a burden of rare mutations in the *APOA5* gene may explain 0.16–0.34% of the heritability. For *LDLR*, for all values, variance explained is between 0.13% and 0.32% of the total phenotypic variance (on the liability scale) and 0.26–0.64% of the heritability.

Sample size extrapolations and power calculations for burden of rare variants. We evaluated the sample size that is needed to reach genome-wide significance levels ($P = 2.5 \times 10^{-6}$) for the T1 test. Our calculations relied on the following assumptions. We assumed that all allelic variants with population frequency less than 1% are causal and have identical effect sizes. We also assumed that all alleles with frequency greater than 1% were benign.

Our calculations differentiate between the allele frequency of a SNV in our exome samples with its true allele frequency in a population. The T1 test compares the number of carriers of an allele for a SNV with sample (rather than population) allele frequency less than 1% among cases and controls. We considered three factors when extrapolating to larger sample sizes. First, we assumed our sample is comprised of 50% cases and 50% controls. As the prevalence of EOMI is estimated to be 5%, the sample frequency of causal alleles is likely to be higher than the population frequency. Second, some alleles with population frequency below 1% may, by chance, have sampling frequency greater than 1% and therefore be excluded from the test. Third, the true allele frequency of the SNVs in the population is unknown. In contrast to earlier work that relied on population genetics modelling[48], we provide an update on the power needed to detect rare variant signal after considering the three factors above. We calculated liberal and conservative estimates for our sample size extrapolations and power calculations. The conservative estimate was based on the estimate of the total population frequency of all causal alleles (below 1%) that would be unlikely to be excluded from the T1 test due to the sampling frequencies exceeding 1%. Because allele frequency distribution is dominated by rare alleles, for an allele with population frequency $\hat{x}$, expected population allele frequency is smaller than x.

$$(E(x|\hat{x}) < \hat{x}) \qquad (1)$$

Therefore, the expected total population frequency of all alleles below frequency x is smaller than the total sampling frequency of alleles below sampling frequency $\hat{x}$. However, setting $\hat{x}$ at 1% would result in a liberal rather than conservative estimate because alleles with population frequency below 1% may be excluded from the T1 test as having sampling frequency above 1%. This occurs due to oversampling cases (our sample has 50% of cases at disease prevalence of 5%) and sampling variance. For example, assuming only one causal allele per gene, the power of the T1 test is maximal for the population allele frequency close to 0.5% for a sample of 1,000 cases and 1,000 controls. For a sample of 10,000 individuals, the chance that a risk allele with population frequency of 0.5% would be excluded from the T1 test is below 10^{-3}, making this threshold even more conservative. Therefore, for a conservative estimate, we have assumed that the total population frequency of all causal alleles per gene would equal the total sampling frequency of alleles below 0.5% in the ESP sample. Our liberal estimate assumed that all causal alleles will be included in the T1 test. We assumed that the total population frequency of all causal alleles per gene would equal the total sampling frequency of alleles below 1% in the ESP sample.

Once we extrapolated the number of mutation carriers to 20,000 samples, we then performed power calculations to see how many samples would be needed to reach a genome-wide significance level for the T1 test ($P = 2.5 \times 10^{-6}$ after correcting for 20,000 genes). Power calculations were performed by first sampling a genotype at random from the pool of 20,000 simulated samples. Based on the T1 carrier status of the drawn sample, we simulated the phenotype based on a calculated probability. The phenotype was simulated based on a prevalence rate of 5% for disease, carrier status of the random sample and assumed relative risk of 2.0 of the mutation. For T1 carriers, the probability of being a case was calculated as relative risk (RR) of T1 carrier multiplied by prevalence rate of disease (RR * prevalence

rate). For non-carriers, the probability of being a case was simply the prevalence rate. The case-control ratio was 1:1. We performed sample size extrapolations for genes with varying number of T1 mutations (25th percentile, median and 75th percentile of carriers with a T1 mutation for all genes discovered in the exome, Supplementary Figs 29–31).

29. Tennessen, J. A. et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337, 64–69 (2012).
30. Antman, E. et al. Myocardial infarction redefined—a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. J. Am. Coll. Cardiol. 36, 959–969 (2000).
31. Fisher, S. et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. 12, R1 (2011).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
33. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genet. 43, 491–498 (2011).
34. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
35. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w (1118); iso-2; iso-3. Fly 6, 80–92 (2012).
36. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
37. Sunyaev, S. et al. Prediction of deleterious human alleles. Hum. Mol. Genet. 10, 591–597 (2001).
38. 1000 Genomes Project Consortium et al. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).
39. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
40. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature Genet. 39, 906–913 (2007).
41. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genet. 38, 904–909 (2006).
42. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
43. Stephens, M., Sloan, J. S., Robertson, P. D., Scheet, P. & Nickerson, D. A. Automating sequence-based detection and genotyping of SNPs from diploid samples. Nature Genet. 38, 375–381 (2006).
44. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202 (1998).
45. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).
46. Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. 91, 839–848 (2012).
47. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. Nature Genet. 42, 348–354 (2010).
48. Kryukov, G. V., Shpunt, A., Stamatoyannopoulos, J. A. & Sunyaev, S. R. Power of deep, all-exon resequencing for discovery of human trait genes. Proc. Natl Acad. Sci. USA 106, 3871–3876 (2009).

LETTER

MicroRNA silencing for cancer therapy targeted to the tumour microenvironment

Christopher J. Cheng[1,2,3]†, Raman Bahal[4], Imran A. Babar[1]†, Zachary Pincus[1]†, Francisco Barrera[3]†, Connie Liu[1], Alexander Svoronos[3], Demetrios T. Braddock[5], Peter M. Glazer[4], Donald M. Engelman[3], W. Mark Saltzman[2] & Frank J. Slack[1]†

MicroRNAs are short non-coding RNAs expressed in different tissue and cell types that suppress the expression of target genes. As such, microRNAs are critical cogs in numerous biological processes[1,2], and dysregulated microRNA expression is correlated with many human diseases. Certain microRNAs, called oncomiRs, play a causal role in the onset and maintenance of cancer when overexpressed. Tumours that depend on these microRNAs are said to display oncomiR addiction[3–5]. Some of the most effective anticancer therapies target oncogenes such as *EGFR* and *HER2*; similarly, inhibition of oncomiRs using antisense oligomers (that is, antimiRs) is an evolving therapeutic strategy[6,7]. However, the *in vivo* efficacy of current antimiR technologies is hindered by physiological and cellular barriers to delivery into targeted cells[8]. Here we introduce a novel antimiR delivery platform that targets the acidic tumour microenvironment, evades systemic clearance by the liver, and facilitates cell entry via a non-endocytic pathway. We find that the attachment of peptide nucleic acid antimiRs to a peptide with a low pH-induced transmembrane structure (pHLIP) produces a novel construct that could target the tumour microenvironment, transport antimiRs across plasma membranes under acidic conditions such as those found in solid tumours (pH approximately 6), and effectively inhibit the miR-155 oncomiR in a mouse model of lymphoma. This study introduces a new model for using antimiRs as anti-cancer drugs, which can have broad impacts on the field of targeted drug delivery.

Silencing aberrantly expressed microRNAs (miRNAs) *in vivo* has been achieved using antisense with various nucleic acid analogues involving locked nucleic acids (LNAs), 2′-*O*-methyl oligonucleotides (for example, antagomiRs), and peptide nucleic acids (PNAs) or nanoencapsulated PNAs[5,9,10]. As with most RNA-based therapies, each of these strategies is stymied by non-specific organ biodistribution, reticuloendothelial system clearance, and endolysosomal trafficking[8,11]. Acidosis is a hallmark of tumours[12]. The pHLIP peptide forms an inducible transmembrane α-helix under acidic conditions[13], has the ability to translocate membrane-impermeable molecules into cells via a non-endocytic route[13,14], and, when administered systemically, can target a variety of epithelial tumours[15]. Exploiting acidity as a general property of the tumour microenvironment, we find that the pHLIP peptide can localize to tumours of lymphoid origin in a subcutaneous flank model (Fig. 1a) and a model of disseminated lymphadenopathy (Fig. 1b), while avoiding the liver. Although pHLIP also shows kidney targeting, much of the peptide is cleared by renal excretion (Extended Data Fig. 1). To exploit these targeting and delivery properties we developed a tumour-targeted antimiR delivery vector (pHLIP-antimiR).

PNAs are nucleic acid analogues comprising nucleobases joined by intramolecular amide bonds. This backbone imparts stability, nuclease resistance, and an increased binding affinity for complementary nucleic acids[16]. We hypothesized that pHLIP would facilitate the intracellular delivery of charge-neutral PNA antimiRs (Fig. 1c), which lack anionic

phosphodiester groups, to cells within the tumour microenvironment. Tethering PNA antimiRs to pHLIP represents a unique approach because the multifunctional peptide component both targets tumours and mediates lipid membrane translocation[13].

Fabrication of pHLIP-antimiR was verified by reversed-phase high-performance liquid chromatography (RP-HPLC), tricine SDS–polyacrylamide gel electrophoresis (SDS–PAGE), electrophoretic mobility shift assay (EMSA), and mass spectrometry (Extended Data Fig. 2a–c). In our constructs, the linkage between the PNA and peptide comprised a disulphide bond, which can be cleaved in the reducing environment of the cytosol (Fig. 1c)[17]; therefore, attachment to the inserting carboxy (C) terminus of pHLIP promotes the intracellular delivery of the PNA antimiR. When administered to A549 cells (Fig. 2a and Extended Data Fig. 2d) and Toledo diffuse large-B cell lymphoma (DLBCL) cells (Extended Data Fig. 2e, f), which express elevated levels of miR-155 compared with other DLBCL cells[18], a pHLIP-antimiR modified with a 5-carboxytetramethylrhodamine (TAMRA) label attached to the PNA resulted in enhanced cellular delivery at acidic extracellular pH compared with neutral pH. PNA delivery to cells by pHLIP does not appear to be greatly affected by sequence since antimiR uptake has been demonstrated with numerous miRNAs including miR-182 (Fig. 2a and Extended Data Fig. 2d), miR-155 (Extended Data Fig. 2e, f), scrambled miR-155, miR-21, and
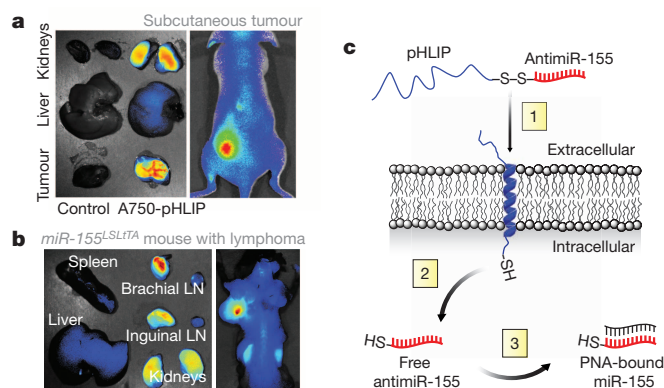

Figure 1 | Targeting miR-155-addicted lymphoma using pHLIP.
a, b, Targeting distribution of pHLIP labelled with Alexa Fluor 750 (A750-pHLIP) 36 h after systemic administration to (**a**) a nude mouse with miR-155 flank tumours (*n* = 3) and (**b**) a *mir-155[LSLtTA]* mouse with lymphadenopathy (*n* = 3); Alexa Fluor 750 conjugated to cysteine was the control. LN, lymph nodes. **c,** Schematic of pHLIP-mediated PNA antimiR delivery. (1) At pH less than 7, the C terminus of pHLIP inserts across lipid bilayers, which facilitates delivery of attached antimiR-155. (2) The disulphide between pHLIP and antimiR-155 is reduced in the cytosol. (3) Intracellular antimiR-155 is free to inhibit miR-155.

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA. [2]Department of Biomedical Engineering, Yale University, New Haven, Connecticut 06511, USA. [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA. [4]Department of Therapeutic Radiology, Yale University, New Haven, Connecticut 06511, USA. [5]Department of Pathology, Yale University, New Haven, Connecticut 06511, USA. †Present addresses: Alexion Pharmaceuticals, Inc., 352 Knotter Drive, Cheshire, Connecticut 06410, USA (C.J.C.); OrbiMed Advisors LLC, 601 Lexington Avenue, 54th Floor, New York, New York 10022, USA (I.A.B.); Departments of Developmental Biology and Genetics, Washington University, St. Louis, Missouri 63110, USA (Z.P.); Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee–Knoxville, Knoxville, Tennessee 37996, USA (F.B.); Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02215, USA (F.J.S.).
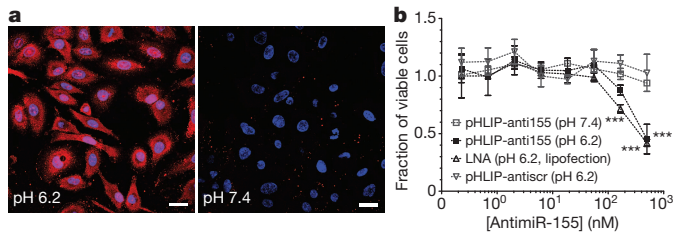
Figure 2 | Intracellular translocation of PNA antimiRs mediated by pHLIP.
a, Confocal projections of A549 cells incubated with labelled pHLIP-antimiR (against control miR-182); scale bars, 25 μm. Red, PNA–TAMRA; blue, nucleus. **b**, Effects of miR-155 inhibition on KB cell viability; all data are normalized to cells treated with vehicle buffer. Data are shown as mean ± s.d., with $n = 3$; statistical analysis performed with two-way analysis of variance (ANOVA); ***$P < 0.001$.

miR-210. Delivery of antimiR-155 by pHLIP (pHLIP-anti155) de-repressed luciferase in miR-155-overexpressing[19] KB cells that stably expressed a miR-155-targeted dual luciferase sensor (Extended Data Fig. 2g). Additionally, inhibition of miR-155 by pHLIP-anti155 reduced KB cell viability at a dose comparable to LNA (15-base oligonucleotide, Exiqon) antimiR-155 delivered by lipofection (Fig. 2b). To demonstrate the adaptability of this antimiR delivery technology to silencing other miRNAs, pHLIP was attached to a PNA antimiR against miR-21, which de-repressed a miR-21 luciferase sensor (Extended Data Fig. 2h). Together, these data suggest that pHLIP-antimiR is effective at delivering PNA antimiRs to multiple cancer cell types, in which endocytosis is hypothesized to be relegated to a supplementary mode of cell uptake due to the transport properties of pHLIP.

Certain oncomiRs have emerged as pharmacological targets. For example, ectopic expression of miR-155 in mice provided the first evidence that dysregulation of a single miRNA could cause cancer[20]. Although aberrant expression of miR-155 is characteristic of numerous cancers, miR-155 is notorious for its oncogenic involvement in lymphomas[21]. We previously developed a Tet-Off-based mouse model in which miR-155 expression is induced in haematological tissues and can be attenuated with the addition of doxycycline (DOX)[5]. Between 2 and 3 months of age, these *mir-155*[LSLtTA] mice develop disseminated lymphoma, in which lymphoid tissues progress from normal histology, to follicular hyperplasia, to follicular lymphoma, to DLBCL (Extended Data Fig. 3a, b). Although these are aggressive cancers comprising neoplastic B cells with a high Ki-67 proliferative index, the disease dramatically regresses upon DOX-induced miR-155 withdrawal (Extended Data Fig. 3b, c). Therefore, this is a model of oncomiR addiction in which tumorigenesis is dependent on expression of miR-155 and its removal leads to cancer regression[22].

We assessed the therapeutic efficacy of pHLIP-anti155 *in vivo* using two tumour models based on *mir-155*[LSLtTA] mice: (1) nude mice subcutaneously implanted with neoplastic B cells derived from the enlarged spleens of *mir-155*[LSLtTA] mice (Extended Data Fig. 4a) and (2) *mir-155*[LSLtTA] mice after progression to conspicuous lymphadenopathy (Extended Data Fig. 4b). Continuous suppression of miR-155 via DOX-impregnated mouse chow or a cocktail of chemotherapeutics and anti-inflammatory steroids (CHOP) served as positive controls that each caused tumour regression (Extended Data Fig. 5a). Since CHOP is part of the current standard of care for human lymphomas[23], the similar response to treatment with DOX and CHOP demonstrated the potential utility of antimiR-155 cancer therapy. Accordingly, intravenous administration of pHLIP-anti155 to the flank tumour model resulted in a significant reduction in tumour growth (Fig. 3a). In a subsequent study at a higher dose, pHLIP-anti155 showed a significant survival advantage compared with a commercially available LNA (Exiqon) antimiR optimized for *in vivo* miR-155 silencing (Fig. 3b and Extended Data Fig. 5b). After administration of pHLIP-anti155, mice exhibited no clinical signs of distress, toxicity, and renal damage (Extended Data Fig. 5c). Note that the dose of pHLIP-anti155 used in this study was much lower (ranging from 17- to 40-fold) than that used in other antimiR delivery reports[10,24].
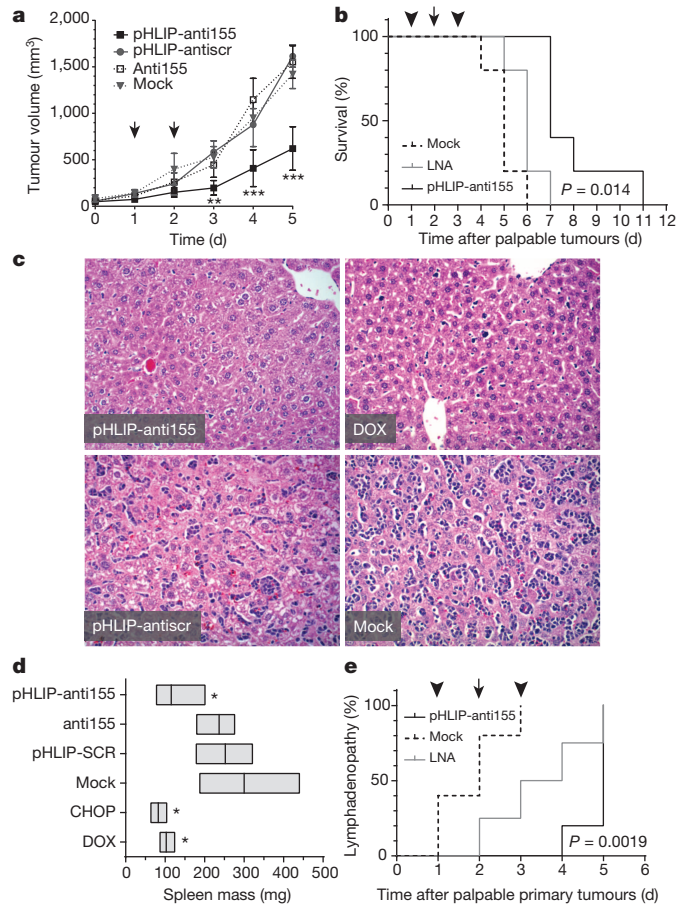

Figure 3 | Targeted silencing of miR-155 has beneficial effects in mice with subcutaneous *mir-155*[LSLtTA] tumours. a, Tumour growth response to treatment; arrows represent 1 mg kg^{-1} PNA dose per intravenous injection; all with $n = 3$, except for pHLIP-anti155 group with $n = 4$. **b**, Survival in response to antimiR treatment; cutoff criteria include tumour volume greater than 1 cm^3 or clinically mandated euthanasia. Symbols represent 2 (arrowhead) or 1 (arrow) mg kg^{-1} intravenous injections; LNA is a fully phosphorothioated LNA antimiR against miR-155; $n = 4$ for all groups; (*) for pHLIP-anti155 compared with LNA. **c**, Representative histological analysis of livers (haematoxylin and eosin (H&E), ×200 magnification) harvested from early endpoint study (Fig. 3a and Extended Data Fig. 5a). **d**, Mass range of spleens from mice in early endpoint study; all with $n = 3$, except for pHLIP-anti155 group with $n = 4$. **e**, Time to development of conspicuous lymphadenopathy in survival study; (**) for pHLIP-anti155 compared with mock. Data are shown as mean ± s.d.; statistical analysis performed with (**a**) two-way ANOVA, (**b, e**) Mantel–Cox test, or (**d**) two-tailed Student's *t*-test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

In addition to delaying tumour growth, pHLIP-anti155 treatment suppressed the metastatic spread of neoplastic lymphocytes to other organs. The liver, lymph nodes, and spleen were common targets for metastatic lymphocytes. In a blinded pathological assessment, livers from mice treated with pHLIP-anti155 and DOX had rare scattered aggregates of one to three neoplastic lymphocytes, while livers in the negative control groups typically had dense tumoral aggregates of up to two dozen cells scattered throughout the entire organ (Fig. 3c)—note that these tissues were harvested at an early endpoint (that is, when the negative controls reached a tumour size of 1 cm^3, Fig. 3a) in relation to the survival study (Fig. 3b) to resolve pharmacological effects. Early endpoint treatment with pHLIP-anti155, DOX, and CHOP reduced the onset of splenomegaly (as judged by spleen mass), which occurred in all of the negative control groups (Fig. 3d). Additionally, pHLIP-anti155 significantly delayed the development of conspicuous lymphadenopathy (Fig. 3e), which was particularly evident in the inguinal and axillary lymph nodes throughout all of the groups (Extended Data Fig. 5d).

On the basis of a blinded complete blood count analysis, the negative control groups comprised a large number of atypical mononuclear cells of lymphoid origin—consistent with the leukaemic phase of lymphoma. Treatment with pHLIP-anti155 and DOX had levels of circulating lymphocytes similar to wild type, while CHOP treatment resulted in lymphocyte levels much lower than wild type (Extended Data Fig. 5e). Although pHLIP can target to metastasized lymph node tumours (Extended Data Fig. 1c), the therapeutic effects on the levels of circulating lymphocytes suggest that the lower incidence of metastatic spread is probably due to antimiR activity at the primary tumour. These findings support the effective targeting of systemic antimiR-155 therapy to neoplastic cells (Extended Data Fig. 5f). The additional lymphopenia caused by CHOP treatment probably reflects the general toxicity of non-targeted conventional chemotherapy drugs (Extended Data Fig. 5e). The absence of systemic toxicity may represent an important advantage for pHLIP-targeted antimiR therapy. Importantly, when healthy C57BL/6 mice were treated at the highest dose and frequency used in this study, pHLIP-anti155 showed no significant impairment of liver and kidney function (Extended Data Fig. 6a). Additionally, white blood cell levels (Extended Data Fig. 6b), body mass (Extended Data Fig. 6c), and organ mass (Extended Data Fig. 6d) were all within normal ranges.

In addition to the miR-155-addicted lymphoma subcutaneous tumour model, pHLIP-anti155 was also effective at treating KB cell xenograft tumours, which stably expressed luciferase for intravital monitoring of tumour bioluminescence (Extended Data Fig. 7), as well as disseminated tumours in *mir-155*[LSLtTA] mice. Although implanted subcutaneous tumour models are effective for evaluating tumour growth, spontaneous cancer models arising in endogenous tissues are a more clinically relevant means of assessing therapeutic efficacy. Remarkably, systemically administered pHLIP-anti155 accumulated in the enlarged lymph nodes of the transgenic *mir-155*[LSLtTA] mice (Fig. 4a). Furthermore, like most therapeutics, PNA oligomers are known to be cleared by the reticuloendothelial system[11], which results in accumulation in the liver; pHLIP-anti155 showed approximately 10-fold reduction in liver accumulation compared with anti155 alone (Fig. 4a and Extended Data Fig. 8a, b). The therapeutic impact of pHLIP-anti155 in *mir-155*[LSLtTA] mice was

supported by a statistically significant decrease in spleen size and a non-statistically significant reduction in lymph node tumour burden (Fig. 4b and Extended Data Fig. 8c–e). A non-statistically significant increase in apoptosis was also observed in the lymph nodes of treated mice (Extended Data Fig. 8f, g). Interestingly, blinded histopathological analysis revealed that spleens in pHLIP-anti155-treated *mir-155*[LSLtTA] mice had differentiated red and white pulp (similar to wild-type mice with no treatment), while the splenic architecture of *mir-155*[LSLtTA] mice treated with pHLIP-antiscr was almost completely effaced (Fig. 4c and Extended Data Fig. 8h). As with the subcutaneous tumour studies, pHLIP-anti155 treatment also showed a 12-fold reduction in liver metastasis (Fig. 4d and Extended Data Fig. 8i, j), while flow cytometric analysis revealed reductions in populations of B220-expressing spleen cells (Extended Data Fig. 8k). Consistent with a lack of systemic toxicity, treatment with pHLIP-anti155 produced no histopathological kidney damage (Extended Data Fig. 8l). Lastly, all mice that developed lymphoma-induced paresis showed improved motor skills after pHLIP-anti155 treatment (Supplementary Videos 1–4, n = 3).

For a more direct assessment of miR-155 silencing in *mir-155*[LSLtTA] mice, we monitored the levels of miR-155 targets in response to antimiR treatment. As an oncogene in lymphoma, miR-155 suppresses genes involved in processes such as apoptosis, proliferation, immune response regulation, as well as cell differentiation and development[21]. However, the addiction mechanisms by which lymphoma regresses upon miR-155 withdrawal are unknown. Typically, miR-155 targets have been identified by differential gene expression analysis of an overexpression condition compared with wild type[25]. To uncover the genes required for miR-155 addiction, we performed RNA sequencing (RNA-seq) analysis on miR-155-addicted lymphoid tumours compared with regressing tumours (Extended Data Fig. 9a). This is the first study, to our knowledge, to identify miRNA cancer targets that directly result from oncomiR withdrawal. Out of 29,209 mouse genes, 2,101 showed significant upregulation or downregulation in response to miR-155 attenuation (Extended Data Fig. 9b and Supplementary Table 1). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of upregulated genes revealed that 41% have been associated with cancer pathways (Extended Data Fig. 9c). Additionally, 25% have been implicated in cell adhesion and migration pathways
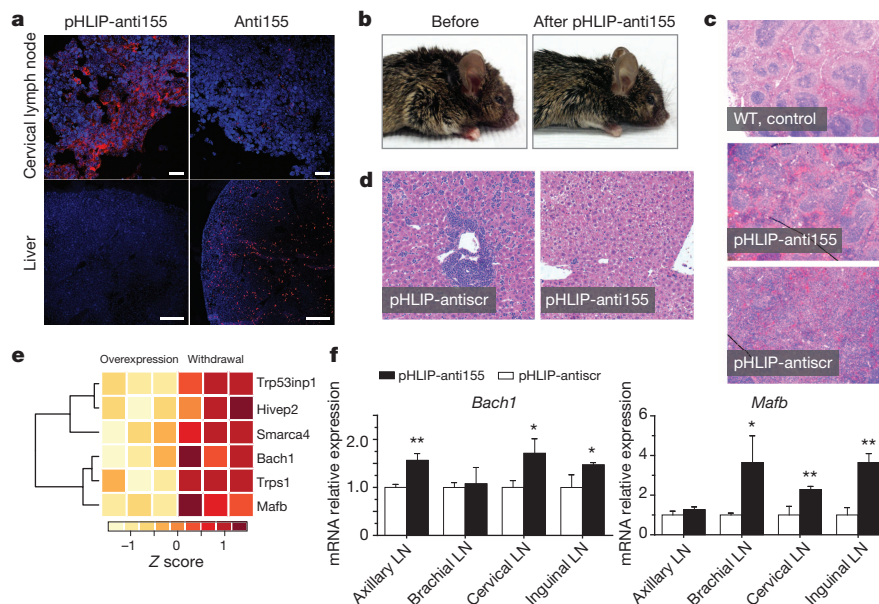

Figure 4 | Delivery of pHLIP-anti155 to *mir-155*[LSLtTA] mice with lymphadenopathy. a, Confocal projections of systemic, tumour-targeted delivery of antimiR-155 to *mir-155*[LSLtTA] mice using pHLIP; scale bars, 25 μm (top, enlarged cervical lymph node) and 250 μm (bottom, liver), n = 3. Red, PNA–TAMRA; blue, nucleus. **b**, Representative *mir-155*[LSLtTA] mouse before and after treatment with pHLIP-anti155, n = 6. **c, d**, Representative H&E analysis of (**c**) spleens and (**d**) livers harvested from diseased littermate *mir-155*[LSLtTA] mice after treatment (n = 6); control spleen represents wild-type mice with no treatment. **e**, Heat map showing selected upregulated genes upon miR-155 withdrawal. **f**, Quantitative PCR (qPCR) determination of gene expression levels in lymphoid tissue from *mir-155*[LSLtTA] mice. Data are shown as mean ± s.d., n = 3; statistical analysis performed with two-tailed Student's t-test; *P < 0.05; **P < 0.01.

such as leukocyte transendothelial migration. We compared the upregulated genes to known and putative miR-155 targets (Supplementary Table 2) identified using the miRWalk target prediction algorithm[26]. At the intersection of these screens, several genes are known to have tumour suppressor characteristics (Fig. 4e, Extended Data Fig. 9d and Supplementary Table 3). One notable gene is *Bach1*, a transcription factor that has been validated as a miR-155 target in renal cancer and cultured B cells[25,27]. Gene expression analysis was used to validate *Bach1* as a miR-155 target in Toledo cells treated with pHLIP-anti155 (Extended Data Fig. 9e) and in *mir-155^LSLtTA* mice undergoing DOX-induced miR-155 withdrawal (Extended Data Fig. 10). Furthermore, diseased *mir-155^LSLtTA* mice treated with pHLIP-anti155 showed an increase in *Bach1* levels in cancerous axillary, cervical, and inguinal lymph nodes (Fig. 4f). A known miR-155 target in lymphoma, *Mafb*[24], was also upregulated in response to pHLIP-anti155 treatment (Fig. 4f). Therefore, pHLIP-anti155 can target lymph node neoplasms and cause effective blockage of miR-155 activity.

While oncomiRs are proving to be potent anticancer targets, in theory, using this approach, every miRNA is a 'druggable' target. Through targeted antagonism of miRNAs, pHLIP-antimiR has vast therapeutic potential for cancer and many other pathological conditions that produce localized acidic environments such as ischaemia, myocardial infarcts, stroke, tissue trauma, and sites of inflammation and infection. The main limitation of this transmembrane delivery approach involves the need for the drug cargo to have limited charge, such as PNA antimiRs. While other antimiR delivery and targeting strategies have been described[28,29], utilization of pHLIP to target the acidic tumour microenvironment is a widely applicable technology that will present new therapeutic and mechanistic opportunities for effective targeting of miRNA silencing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
2. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nature Rev. Cancer* **6**, 857–866 (2006).
3. Esquela-Kerscher, A. & Slack, F. J. Oncomirs — microRNAs with a role in cancer. *Nature Rev. Cancer* **6**, 259–269 (2006).
4. Medina, P. P., Nolde, M. & Slack, F. J. OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **467**, 86–90 (2010).
5. Babar, I. A. *et al.* Nanoparticle-based therapy in an *in vivo* microRNA-155 (miR-155)-dependent mouse model of lymphoma. *Proc. Natl Acad. Sci. USA* **109**, E1695–E1704 (2012).
6. Stenvang, J., Petri, A., Lindow, M., Obad, S. & Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* **3**, 1 (2012).
7. Kasinski, A. L. & Slack, F. J. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nature Rev. Cancer* **11**, 849–864 (2011).
8. Cheng, C. J., Saltzman, W. M. & Slack, F. J. Canonical and non-canonical barriers facing antimiR cancer therapeutics. *Curr. Med. Chem.* **20**, 3582–3593 (2013).
9. Elmen, J. *et al.* LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896–899 (2008).
10. Krutzfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
11. White, P. J., Anastasopoulos, F., Pouton, C. W. & Boyd, B. J. Overcoming biological barriers to *in vivo* efficacy of antisense oligonucleotides. *Expert Rev. Mol. Med.* **11**, e10 (2009).
12. Vaupel, P., Kallinowski, F. & Okunieff, P. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. *Cancer Res.* **49**, 6449–6465 (1989).
13. Reshetnyak, Y. K., Andreev, O. A., Lehnert, U. & Engelman, D. M. Translocation of molecules into cells by pH-dependent insertion of a transmembrane helix. *Proc. Natl Acad. Sci. USA* **103**, 6460–6465 (2006).
14. Thevenin, D., An, M. & Engelman, D. M. pHLIP-mediated translocation of membrane impermeable molecules into cells. *Chem. Biol.* **16**, 754–762 (2009).
15. Reshetnyak, Y. K. *et al.* Measuring tumor aggressiveness and targeting metastatic lesions with fluorescent pHLIP. *Mol. Imaging Biol.* **13**, 1146–1156 (2011).
16. Nielsen, P. E., Egholm, M. & Buchardt, O. Peptide nucleic acid (PNA). A DNA mimic with a peptide backbone. *Bioconjug. Chem.* **5**, 3–7 (1994).
17. Østergaard, H., Tachibana, C. & Winther, J. R. Monitoring disulfide bond formation in the eukaryotic cytosol. *J. Cell Biol.* **166**, 337–345 (2004).
18. Rai, D., Karanti, S., Jung, I., Dahia, P. L. M. & Aguiar, R. C. T. Coordinated expression of microRNA-155 and predicted target genes in diffuse large B-cell lymphoma. *Cancer Genet. Cytogenet.* **181**, 8–15 (2008).
19. Rather, M. I., Nagashri, M. N., Swamy, S. S., Gopinath, K. S. & Kumar, A. Oncogenic microRNA-155 down-regulates tumor suppressor CDC73 and promotes oral squamous cell carcinoma cell proliferation: implications for cancer therapeutics. *J. Biol. Chem.* **288**, 608–618 (2013).
20. Costinean, S. *et al.* Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in Eμ-miR155 transgenic mice. *Proc. Natl Acad. Sci. USA* **103**, 7024–7029 (2006).
21. Sandhu, S. K., Croce, C. M. & Garzon, R. Micro-RNA expression and function in lymphomas. *Adv. Hematol.* **2011**, 1–12 (2011).
22. Cheng, C. J. & Slack, F. J. The duality of oncomiR addiction in the maintenance and treatment of cancer. *Cancer J.* **18**, 232–237 (2012).
23. DeVita, V. T., Lawrence, T. S. & Rosenberg, S. A. *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology* 9th edn (Lippincott Williams & Wilkins, 2011).
24. Zhang, Y. *et al.* LNA-mediated anti-miR-155 silencing in low-grade B-cell lymphomas. *Blood* **120**, 1678–1686 (2012).
25. Gottwein, E. *et al.* A viral microRNA functions as an orthologue of cellular miR-155. *Nature* **450**, 1096–1099 (2007).
26. Dweep, H., Sticht, C., Pandey, P. & Gretz, N. miRWalk–database: prediction of possible miRNA binding sites by 'walking' the genes of three genomes. *J. Biomed. Inform.* **44**, 839–847 (2011).
27. Li, S. *et al.* microRNA-155 silencing inhibits proliferation and migration and induces apoptosis by upregulating BACH1 in renal cancer cells. *Mol. Med. Report.* **5**, 949–954 (2012).
28. Brognara, E. *et al.* Uptake by human glioma cell lines and biological effects of a peptide-nucleic acids targeting miR-221. *J. Neurooncol.* **118**, 19–28 (2014).
29. Wang, Y.-Z. *et al.* Delivery of an miR155 inhibitor by anti-CD20 single-chain antibody into B cells reduces the acetylcholine receptor-specific autoantibodies and ameliorates experimental autoimmune myasthenia gravis. *Clin. Exp. Immunol.* **176**, 207–221 (2014).

METHODS

PNA synthesis. Regular Boc-protected PNA monomers were purchased from ASM Research Chemicals. All the given oligomers were synthesized on solid-support using standard Boc chemistry procedures[30]. The oligomers were cleaved from the resin using *m*-cresol:thioanisole:TFMSA:TFA (1:1:2:6) cocktail solution. The resulting mixtures were precipitated with ether (three times), purified, and characterized by RP-HPLC and MALDI-TOF, respectively. All PNA stock solutions were prepared using nanopure water and the concentrations were determined at 90 °C on a Cary 3 Bio spectrophotometer using the following extinction coefficients: 13,700 M^{-1} cm^{-1} (A), 6,600 M^{-1} cm^{-1} (C), 11,700 M^{-1} cm^{-1} (G), and 8,600 M^{-1} cm^{-1} (T). The 23-base oligonucleotide PNA oligomer complementary to miR-155 has an estimated melting temperature (T_m) of 77.8 °C. Single-isomer 5-carboxytetramethylrhodamine (TAMRA) purchased from VWR was exclusively conjugated to the amino (N) terminus of PNAs with a hydrophilic bifunctional linker, Boc-miniPEG-3^{Tm} (11-amino-3,6,9-trioxaundecanoic acid, DCHA, denoted in the sequences by -ooo-) purchased from Peptide International. Cysteine was also conjugated to the C terminus of PNAs using a Boc-miniPEG-3 linker.

The following PNA antimiR sequences were used: anti155, TAMRA–ooo-ACC CCTATCACAATTAGCATTAA-ooo-Cys; antiscr, TAMRA–ooo-ACCCAATC GTCAAATTCCAATAA-ooo-Cys; anti21, TAMRA–ooo-TCAACATCAGTCTG ATAAGCTA-ooo-Cys; anti182, TAMRA–ooo-CGGTGTGAGTTCTACCATTG CCAAA-ooo-Cys.

Full-length PNA antimiRs were used throughout this study. While current technologies such as 'tiny' LNAs have seen efficacy with miRNA seed-targeted 8-base oligonucleotide antimiRs[31], truncated PNA antimiRs should be similarly effective owing to their high binding affinity, which can be further enhanced with chemical modifications[32].

Synthesis and characterization of pHLIP-antimiR. To generate pHLIP-antimiR constructs, the following pHLIP sequence (New England Peptide) was synthesized: AAEQNPIYWARYADWLFTTPLLLLDLALLVDADEGT(CNPys)G; conjugation of the C terminus to thiolated-PNA was facilitated by incorporating a cysteine group derivatized with 3-nitro-2-pyridinesulphenyl (NPys). To synthesize pHLIP-antimiR constructs, pHLIP-Cys(NPys) and antimiR PNA (peptide:PNA 1:1.3) were reacted overnight in the dark in a mixture of DMSO/DMF/0.1 mM KH_2PO_4 pH 4.5 (v/v 3:1:1) under argon. Note that this protocol was adapted from a general method of conjugating peptides to PNAs. Aside from pHLIP, attaching molecules, such as cell-penetrating peptides, to PNAs can increase cellular uptake and *in vivo* delivery efficacy[33,34]. However, these conjugates typically require high doses and distribute to tissues throughout the body, which can result in off-target effects[11,35]. Similarly, pHLIP can be attached to other antimiR compositions (such as LNA), which would probably improve tumour targeting; however, physicochemical properties of PNA make them more amenable to pHLIP-mediated membrane translocation. A750-pHLIP was fabricated as previously described[15].

Purification and verification of pHLIP-antimiR. After conjugation, pHLIP-antimiR was purified by RP-HPLC (Shimadzu) using a C18 column and a mobile phase gradient of water and acetonitrile with 0.1% trifluoroacetic acid. Purified pHLIP-antimiR was further characterized using matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF). Concentrations of pHLIP-antimiR were determined on a Nanodrop Spectrophotometer (Thermo Scientific) at 260 nm corrected for peptide and TAMRA absorbance. Gelshift analysis used a 20% TBE gel and Bolt electrophoresis system; before loading, samples were incubated with an equimolar amount of miR-155, denatured at 95 °C for 2 min, and allowed to anneal at 37 °C for 30 min. SYBR Gold (Life Technologies) was used to visualize miR-155; pHLIP and free PNA were not detected by the stain. Tricine SDS–PAGE used a 16% tricine gel (Life Technologies) and standard SDS–PAGE procedures. Samples were visualized first using TAMRA fluorescence on a Maestro 2 Multispectral Imaging System (PerkinElmer), and then using Simply Blue Coomassie stain (Life Technologies). For disulphide reduction studies, pHLIP-antimiR was reduced for 30 min in 200 mM DTT for HPLC and EMSA, and 5 mM TCEP for tricine SDS–PAGE. For all *in vitro* and *in vivo* studies, pHLIP-antimiR was heated at 65 °C for 10 min to prevent aggregation.

Animals. All mice were maintained at Yale University in accordance with Yale Animal Resource Center and the Institutional Animal Care and Use Committee guidelines. The *mir-155LSLtTA* mice were generated as previously described[5]. For transplant studies, 5- to 6-week-old female CrTac:NCr-Foxn1 nude mice (Taconic) were used. For toxicology studies, 8- to 9-week-old female C57BL/6J mice (Jackson) were used. For treatment of *mir-155LSLtTA* mice, a sample size of at least four was appropriate on the basis of *post hoc* power analysis using quantitation of spleen size (Extended Data Fig. 8d) with a 95% confidence interval. For all animal studies, group allocations were randomized and all pathological analyses were blinded to treatment groups and expected experimental outcomes.

Cell culture. For all pH-controlled cell culture experiments, cells (previously tested for mycoplasma and supplied from ATCC) were incubated with 10% FBS in RPMI

buffered at pH 7.4 with HEPES or pH 6.2 with MES, and treated with pHLIP-antimiR suspended in reaction buffer which constituted no more than 1% of the final volume.

Histology and other techniques. Harvested tissues were fixed in 10% formalin and processed by Yale Pathology Tissue Services for H&E and terminal deoxynucleotidyl-transferase-mediated dUTP nick end labelling (TUNEL) staining. Retro-orbitally collected whole blood preserved in EDTA or serum separated using lithium heparin was sent to Antech Diagnostics for complete blood count or clinical chemistry analyses, respectively. Image quantification used ImageJ version 1.47 (NIH) and Colour Deconvolution plugin (A. C. Ruifrok). Intravital and *ex vivo* fluorescence imaging was performed on either an IVIS Spectrum System (Caliper) or Maestro 2 Multispectral Imaging System using near-infrared or TAMRA filter sets. Live mice were anaesthetized using isoflurane during image acquisition. For whole-organ studies, organs were harvested and fixed in 10% formalin before imaging.

Flank tumour establishment. To establish *mir-155LSLtTA* lymphoma subcutaneous flank tumours, first enlarged spleens were extracted from 2- to 3-month-old *mir-155LSLtTA* mice with obvious lymphadenopathy (which generally correlated with incidence of splenomegaly). Using 100 μm pore-size cell strainer technique, spleen tissue was dispersed into a single cell suspension in 5% FBS in PBS on ice. Red blood cells were lysed using ammonium chloride lysis buffer (Stem Cell Technologies), and 5×10^6 cells were subcutaneously injected into nude mice. Tumours were generally palpable within 10 days; tumour volume was calculated as (length × width^2)/2.

For bioluminescent xenograft tumours, KB cells were stably transfected with firefly luciferase and clonally selected via hygromycin B selection; 5×10^6 cells were subcutaneously injected into nude mice to establish tumours. RediJect D-Luciferin Ultra Bioluminescent Substrate (PerkinElmer) was administered via the manufacturer's protocol for intravital monitoring of tumour bioluminescence using IVIS Spectrum (Caliper). It was pre-established that, for all flank tumour studies, animals were excluded if their tumours had not reached a volume of 50–100 cm^3 by the time of treatment. Animals were randomized into experimental arms by minimizing the differences in mean tumour size and standard deviation.

Confocal imaging and flow cytometry. For fixed cell confocal preparation, after treatment for 1 h at with 5 μM of pHLIP-anti155 (Fig. 2a), cells were washed with 1% BSA in PBS, fixed in 4% paraformaldehyde, and permeabilized using 0.1% Triton X-100 in PBS. All washes were performed using PBS at pH 7.4 to wash away surface-bound pHLIP. Actin and nuclei were stained with Texas Red-X phalloidin (Life Technologies) and Hoechst 33342 (Life Technologies), respectively. Cells were mounted in Slow Fade Gold (Invitrogen). Alternatively, Toledo cells were treated with 500 nM pHLIP-anti155 (Extended Data Fig. 2e), washed with 1% BSA in PBS, and imaged live without fixation or permeabilization. For tumour and liver tissues, organs were harvested and fixed in 10% formalin, then incubated overnight in 30% sucrose in PBS. Tumours were flash frozen in OCT before being slicing into 10 μm sections, permeabilized, stained, and mounted in Vectashield (Vector Labs). Cell and tissue confocal imaging used a TCS SP5 Spectral Confocal Microscope (Leica); confocal projections were constructed using LAS AF software (Leica) with 0.9-μm-stack height. For live cell flow cytometry, after 48 h of treatment, cells were washed five times with 1% BSA in PBS on ice and then analysed on a FACScan (BD Biosciences) using FlowJo software (Tree Star); for B220 studies, freshly harvested spleen cells (see section on Flank tumour establishment) were blocked with 10% FBS (20 min), stained with Alexa488-anti-CD45R/B220 (BD clone RA3-6B2, 20 min incubation at room temperature at 1 μg ml^{-1} concentration), washed three times with PBS on ice, and transferred to 1% BSA 0.1% NaN_3 in PBS on ice before analysis.

Luciferase reporter and cell viability. For dual luciferase reporter experiments, the miRNA sensor was generated by inserting the target sequence for miR-155 into the 3′ untranslated region of *Renilla* luciferase on a psiCHECK-2 vector (Promega). KB cells were stably transfected using Lipofectamine 2000 (Life Technologies) and co-transfection with a Linear Hygromycin Marker (Clontech) followed by clonal selection. Utilization of stable clones was more reliable than transiently transfected cells for antimiR studies. Cell lysates were measured for luciferase activity 48 h after treatment using the Dual-Luciferase Reporter Assay System (Promega). Control LNA antimiR-155 (Exiqon) was delivered by lipofectamine RNAiMAX. Optimal sensor activity was seen at a 500 nM dose, although inhibition of miR-155 was also observed at lower doses. For analysis of miR-21 inhibition A549 cells were similarly treated with pHLIP-anti21 and relevant controls; however, the cells were instead transfected with a miR-21-specific LightSwitch miRNA Target GoClone Luciferase Reporter (Active Motif). Cell viability was measured 96 h after treatment using CellTiter-Glo (Promega). For both luciferase and viability assays, all treatments were performed at the indicated pH for 24 h, then media was replaced with 10% FBS in RPMI at physiological pH for extended incubation.

qPCR. For qPCR analysis of tissue after treatment with two 2 mg kg^{-1} injections of pHLIP-anti155 or pHLIP-antiscr spaced 48 h apart, tissues were harvested 24 h after the last injection and divided into at least five representative 1 mg slices. Tissue slices were pooled into Trizol (Life Technologies) and homogenized using a Precellys 24

Homogenizer. As per the manufacturer's protocol, chloroform was added to facilitate phase separation, and the RNA-containing aqueous phase was collected. An equal volume of 200 proof ethanol was added, and RNA was purified from this mixture using RNeasy kit (Qiagen) and standard procedures with on-column DNase I digestion; standard RNeasy purification was followed for RNA extraction from cells. Reverse transcription PCR was performed with 1 µg total RNA and poly-A based iScript cDNA Synthesis Kit (Bio-Rad). Real-time PCR was performed with Quantitect Primer Assays (Qiagen) and iQ SYBR Green Supermix (Bio-Rad) using a Roche LightCycler 480 System; all samples were normalized to β-actin.

RNA-seq. For RNA-seq analysis, the overexpression and withdrawal groups both consisted of three mice with subcutaneous tumours that were established from enlarged spleens of diseased donor *mir-155^{LSLtTA}* mice (Extended Data Fig. 9a). The overexpression and withdrawal mice were paired such that each of the three pairs was from a separate donor littermate. Tissue was harvested once tumours reached a volume of 1 cm^3; for mice in the miR-155 withdrawal group, DOX was administered for 16 h before tissue collection. As described in the transplant methods, tumour tissue was dispersed into a single cell suspension and red blood cells were lysed. Total RNA was extracted from the remaining cells using the hybrid Trizol and RNeasy protocol described in the qPCR methods. High-quality total RNA (Agilent 2100 Bioanalyzer RIN value greater than 7) was sent to Expression Analysis for library preparation, Illumina TruSeq mRNA sequencing (50-base-pair paired end, 25 million reads per sample), alignment to the mouse genome (greater than 80% aligned to the NCBI37/mm9 assembly), and counts of the number of gene-mapped fragments given the maximum likelihood abundances. DESeq was used first to estimate size factors (that is, normalize samples by their respective sizes) and dispersions (that is, variance between samples), and then to identify differentially expressed genes (Supplementary Table 1). Heat maps were generated using variance stabilizing transformations of the count data on the basis of a parametric fit to the overall mean dispersions.

KEGG analysis. The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov) was used to identify the KEGG pathways that were enriched in the genes and both upregulated in response to miR-155 withdrawal and having a false discovery rate less than 0.05. Enriched KEGG pathways had a minimum count threshold of 2 and a modified Fisher's exact *P* value for gene enrichment less than 0.05.

30. Christensen, L. *et al.* Solid-phase synthesis of peptide nucleic acids. *J. Pept. Sci.* **1,** 175–183 (1995).
31. Obad, S. *et al.* Silencing of microRNA families by seed-targeting tiny LNAs. *Nature Genet.* **43,** 371–378 (2011).
32. Sahu, B. *et al.* Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing γ-peptide nucleic acids with superior hybridization properties and water solubility. *J. Org. Chem.* **76,** 5614–5627 (2011).
33. Fabani, M. M. *et al.* Efficient inhibition of miR-155 function in vivo by peptide nucleic acids. *Nucleic Acids Res.* **38,** 4466–4475 (2010).
34. Shiraishi, T. & Nielsen, P. E. Peptide nucleic acid (PNA) cell penetrating peptide (CPP) conjugates as carriers for cellular delivery of antisense oligomers. *Artif. DNA PNA XNA* **2,** 90–99 (2011).
35. Fabbri, E. *et al.* miRNA therapeutics: delivery and biological activity of peptide nucleic acids targeting miRNAs. *Epigenomics* **3,** 733–745 (2011).
36. Xu, G. *et al.* Transcriptome and targetome analysis in MIR155 expressing cells using RNA-seq. *RNA* **16,** 1610–1622 (2010).
37. Loeb, G. B. *et al.* Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Mol. Cell* **48,** 760–770 (2012).

a

b

c
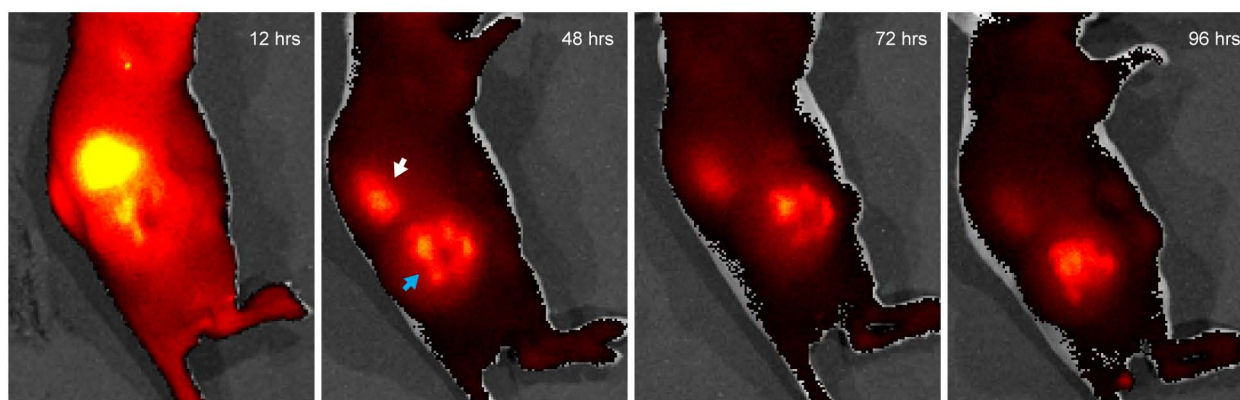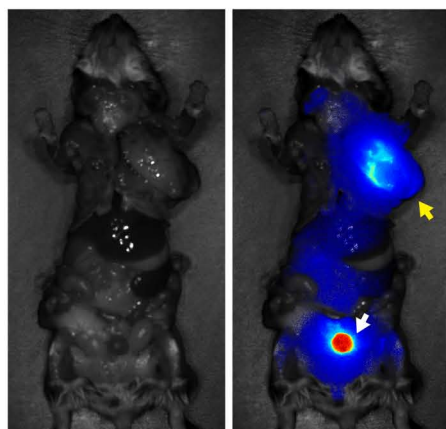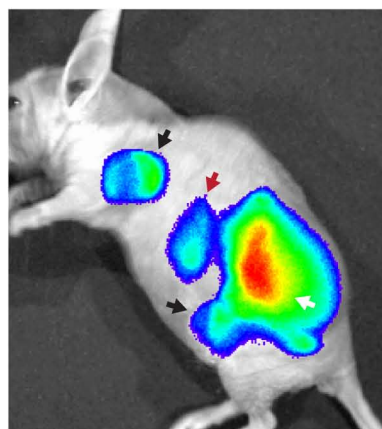
Extended Data Figure 1 | Distribution of pHLIP to the renal system and lymph node metastases. a, Intravenous injection of A750-pHLIP distributes to the (white arrow) kidneys and (blue arrow) tumour in a representative *mir-155*[LSLtTA] subcutaneous flank model (*n* = 3); time points indicate hours after a single injection of A750-pHLIP. Previous reports have observed systemic distribution of pHLIP to kidneys in other mouse models[15]. Similarly, we speculate that the increased uptake of pHLIP peptide in the kidneys is due to excretion and increased acidity of renal tubule cells. Initially kidneys are highly enriched for pHLIP, which is gradually excreted while pHLIP shows a more steady accumulation in the tumour. **b,** Representative example showing A750-pHLIP distribution to the (white arrow) bladder and (yellow arrow) enlarged axillary lymph node 36 h after intravenous administration into *mir-155*[LSLtTA] mice with lymphadenopathy (*n* = 3). **c,** In addition to distributing to the (white arrow) primary *mir-155*[LSLtTA] flank tumour and (red arrow) kidneys, A750-pHLIP distributes to (black arrows) enlarged lymph nodes that resulted from metastatic spread; intravital fluorescence of A750-pHLIP was detected 48 h after intravenous injection into nude transplant mice with conspicuous lymphadenopathy (a representative animal from *n* = 3 is shown).
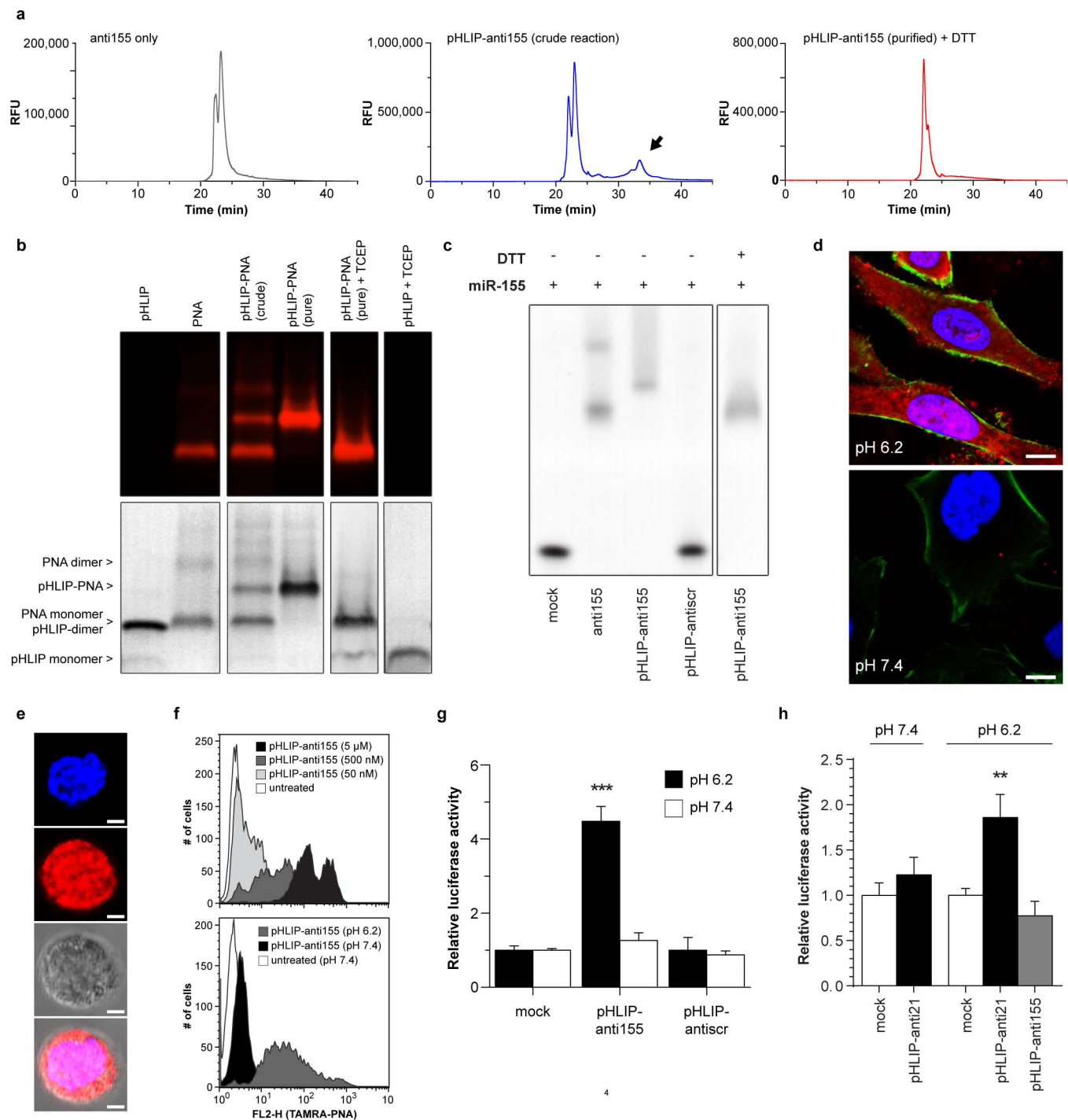
Extended Data Figure 2 | Assessment of pHLIP–PNA conjugation and activity. a, HPLC elution profiles of (top) free PNA, (middle) reaction mixture of PNA and pHLIP-C(Npys), and (bottom) purified pHLIP–PNA incubated in DTT. HPLC was used to purify pHLIP–PNA (black arrow). The fluorescence detection of TAMRA (ex/em: 540/575) that was conjugated to the PNA is shown; samples were also detected by absorbance at 260 and 280 nm (data not shown). **b,** Tricine SDS–PAGE evaluation of pHLIP–PNA conjugation. Gel was visualized by (top) TAMRA fluorescence to detect labelled PNA and (bottom) Coomassie stain to detect both PNA and peptide. **c,** Gelshift analysis of pHLIP-antimiR-155 binding to miR-155 and disulphide reduction in the presence of DTT. **d,** High-magnification confocal projections of A549 cells incubated with labelled pHLIP-antimiR (against control miR-182); scale bars, 7.5 μm. The diffuse intracellular fluorescence is indicative of freely distributed

antimiR throughout the cytosol—note that the presence of marginal punctate fluorescence at both pH levels suggests that endocytosis is probably an additional mode of cell entry. **e,** Toledo DLBCL lymphocytes were incubated with labelled pHLIP-anti155 at pH 6.2; fluorescence of a representative live cell is overlaid on a bright field micrograph; scale bars, 2 μm. **f,** Flow cytometry analysis of Toledo cells incubated with labelled pHLIP-anti155; cell association was dependent on dose (top, pH 6.2) and pH (bottom, 500 nM dose). **g,** Inhibition of miR-155 demonstrated by de-repression of a miR-155 dual luciferase sensor in KB cells. **h,** Inhibition of miR-21 demonstrated by de-suppression of luciferase expression in A549 cells transfected with a *Renilla* luciferase sensor. Data are shown as mean ± s.d., with $n = 3$; statistical analysis performed with two-tailed Student's *t*-test; **$P < 0.01$; ***$P < 0.001$.
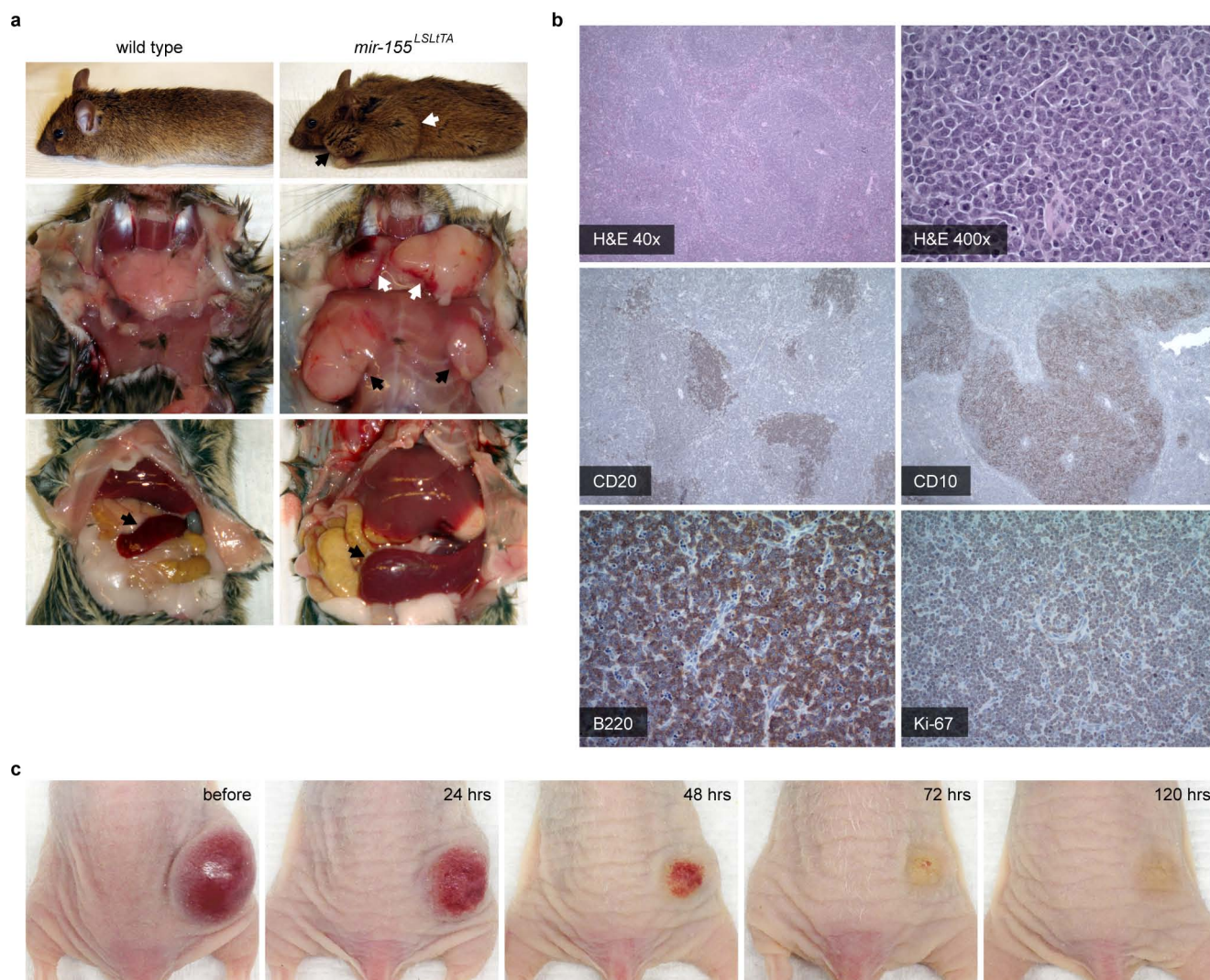
Extended Data Figure 3 | Pathology of the *mir-155^LSLtTA* model of oncomiR addiction. a, Organomegaly in representative diseased *mir-155^LSLtTA* mice: top, conspicuous lymphadenopathy seen in the (black arrow) cervical and (white arrow) brachial lymph nodes; middle, enlarged exposed (white arrows) cervical and (black arrows) axillary lymph nodes; bottom, enlarged (black arrows) spleen. **b**, Histopathology of *mir-155^LSLtTA* mice: H&E stain of an enlarged spleen shows expansion of the white pulp by a nodular, neoplastic infiltrate; staining of the spleen shows CD20$^+$ and CD10$^+$ B cells of follicular centre origin. Analysis of enlarged lymph nodes indicates DLBCL with lymph node architecture effaced by a confluent population of B220$^+$ neoplastic lymphocytes and a Ki-67 proliferative index at nearly 100%; $n = 5$. **c**, Tumour regression due to DOX-induced miR-155 withdrawal in a subcutaneous tumour model established from transplanted splenic *mir-155^LSLtTA* lymphocytes; time points indicate hours after initial administration of DOX. With a cancer phenotype that is relevant to human disease yet can be modulated by miRNA silencing, this is an excellent model for evaluating miR-155-targeted therapies.

a

miR-155-induced lymphoma

n=1

harvest lymphocytes

establish flank tumors

subcutaneous tumor

1. pHLIP-anti155
2. pHLIP-antiscr
3. anti155 (PNA only)
4. mock
5. DOX
7. LNA (against miR-155)
6. CHOP

injections of 2 or 1 mg/kg (IV)

n=3-5

Therapeutic Analysis (endpoint)

average time to palpable tumor 10 days

Days 1-3

Days 5-11

b

miR-155-induced lymphoma

1. pHLIP-anti155
2. pHLIP-antiscr
3. anti155 (PNA only)
4. mock

injection 2 mg/kg (IV)

n=3

Biodistribution (endpoint)

or

injection 2 mg/kg (IV)

n=4

Therapeutic Analysis (endpoint)

average induction time ~2 months

Day 1

Day 3

Day 4

Extended Data Figure 4 | Experimental schematics for mouse tumour studies. a, Workflow for treatment of the *mir-155*^{LSLtTA} subcutaneous flank model for the early endpoint and survival studies; day 1 indicates time of first injection. For the 'early treatment' experiments in Fig. 3a, b, d–f, h and Extended Data Fig. 5b, c, mice were treated on days 1 and 2 with pHLIP-anti155, mock buffer, pHLIP-antiscr and anti155 only; fed DOX starting on day 3; or treated with CHOP regimen on days 2–4. For survival experiments in Fig. 3c, g and Extended Data Fig. 5a, mice were treated on days 1–3 with pHLIP-anti155, LNA against miR-155, and mock buffer. **b**, Workflow for investigation of the *mir-155*^{LSLtTA} model of lymphoma for the biodistribution and miR-155 silencing studies. For experiments in Fig. 4a and Extended Data Fig. 8a, b, mice were treated on day 1 with pHLIP-anti155, anti155 only, and mock buffer. For experiments in Fig. 4b–d, h and Extended Data Fig. 8c–g, mice were treated on day 1 and day 3 with pHLIP-anti155, pHLIP-antiscr, and mock buffer, or fed DOX 16 h before harvest.
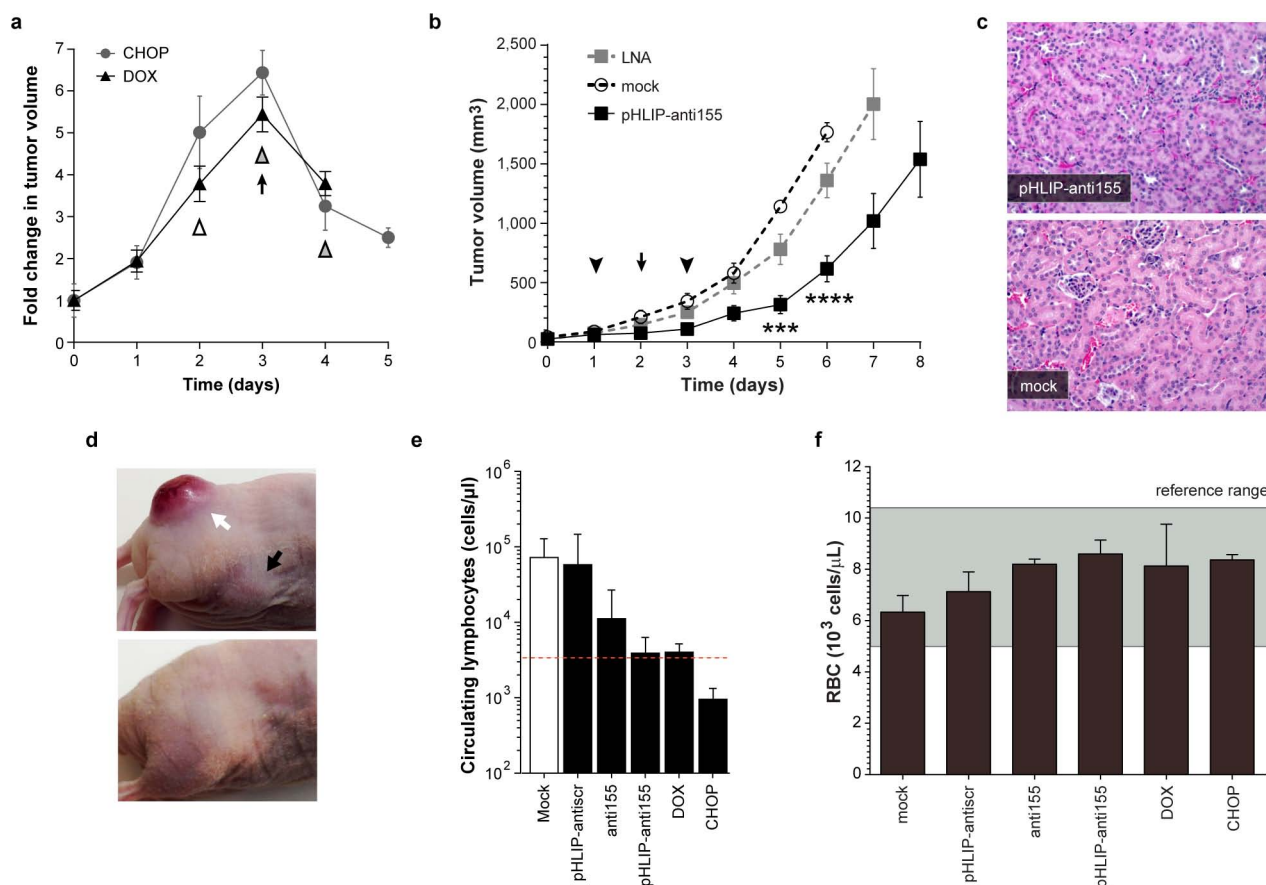
Extended Data Figure 5 | Administration of pHLIP-anti155 to mice with subcutaneous lymphoma flank tumours. a, Fold change in tumour size in response to miR-155 withdrawal and CHOP treatment ($n = 3$); arrow represents initiation of DOX treatment ($n = 3$, food pellets enriched with DOX at 2.3 g/kg, Bio-Serv), white triangle represents CHOP treatment (systemic injection of cyclophosphamide at 40 mg/kg, doxorubicin at 3.3 mg/kg, and vincristine at 0.5 mg/kg; oral gavage of prednisone at 0.2 mg/kg), grey triangles represent maintenance administration of prednisone. **b**, Tumour growth response to systemically administered antimiR treatment; symbols represent intravenous injections of 2 (arrowhead) or 1 (arrow) mg kg^{-1} of pHLIP-conjugated antimiR-155, molar equivalent of phosphorothioated antimiR-155 LNA, or mock delivery solution; $n = 5$, data are shown as mean $\pm$ s.e.m.; statistical comparison of pHLIP-anti155 to LNA performed with two-way ANOVA; ***$P < 0.001$, ****$P < 0.0001$. c, Representative histological analysis of kidneys (H&E, $\times 100$ magnification) harvested from early endpoint study, in which all of the mice from Fig. 3a and Extended Data Fig. 5a were killed at the same time for analysis. Kidney sections reveal an absence of microscopic changes in treated animals (pHLIP-anti155) that would be indicative of renal toxicity (compare with normal renal sections in mock control). **d**, Representative pHLIP-antiscr-treated mouse (top) with primary flank tumour (white arrow) and enlarged inguinal lymph node (black arrow) compared with an untreated mouse with no tumour (bottom). **e**, Measurement of circulating lymphocytes in blood collected at time of death in early endpoint study; dotted line denotes average level in nude mice with no tumour. **f**, Although pHLIP interacts with the outer leaflet of lipid membranes, no significant change in red blood cell (RBC) levels was detected after intravenous treatment of mice with subcutaneous *mir-155^{LSLtTA}* transplant tumours. This supports the specificity of pHLIP treatments on cells of tumour origin since pHLIP-antimiR treatments affect the levels of circulating lymphocytes (Extended Data Fig. 5e); data are shown as mean $\pm$ s.d.
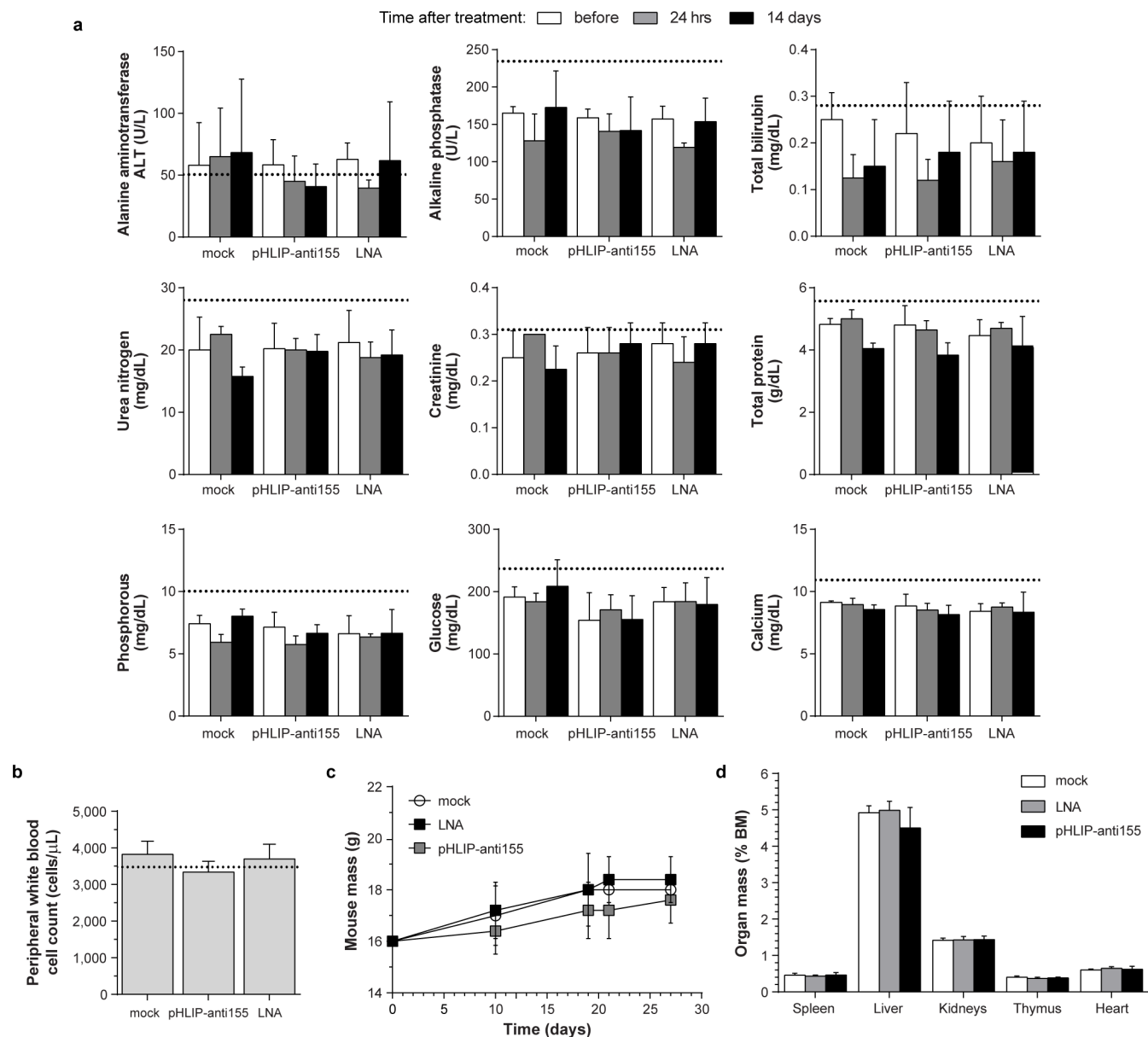
Extended Data Figure 6 | Toxicology assessment of intravenously administered pHLIP-anti155 to C57BL/6J mice. a, Serum-based clinical chemistry evaluation of systemic toxicity with focus on liver and kidney function; dosing schedule consisted of injections of 2 mg kg^{-1} of pHLIP-anti155 (and equimolar dose of LNA) on days 10 and 12, as well as 1 mg kg^{-1} on day 11. Blood samples were serially harvested retro-orbitally on day 0 (10 days before start of treatment), as well as 1 day and 14 days after treatment. **b**, Circulating white blood cell count collected 14 days after treatment. **c**, Mouse mass throughout duration of the study. **d**, Organ mass normalized to total body mass at time of harvest. **a–d**, For all analyses mock $n = 4$, pHLIP-anti155 $n = 5$, and LNA $n = 5$; dotted lines indicate typical wild-type values for C57BL/6J mice.
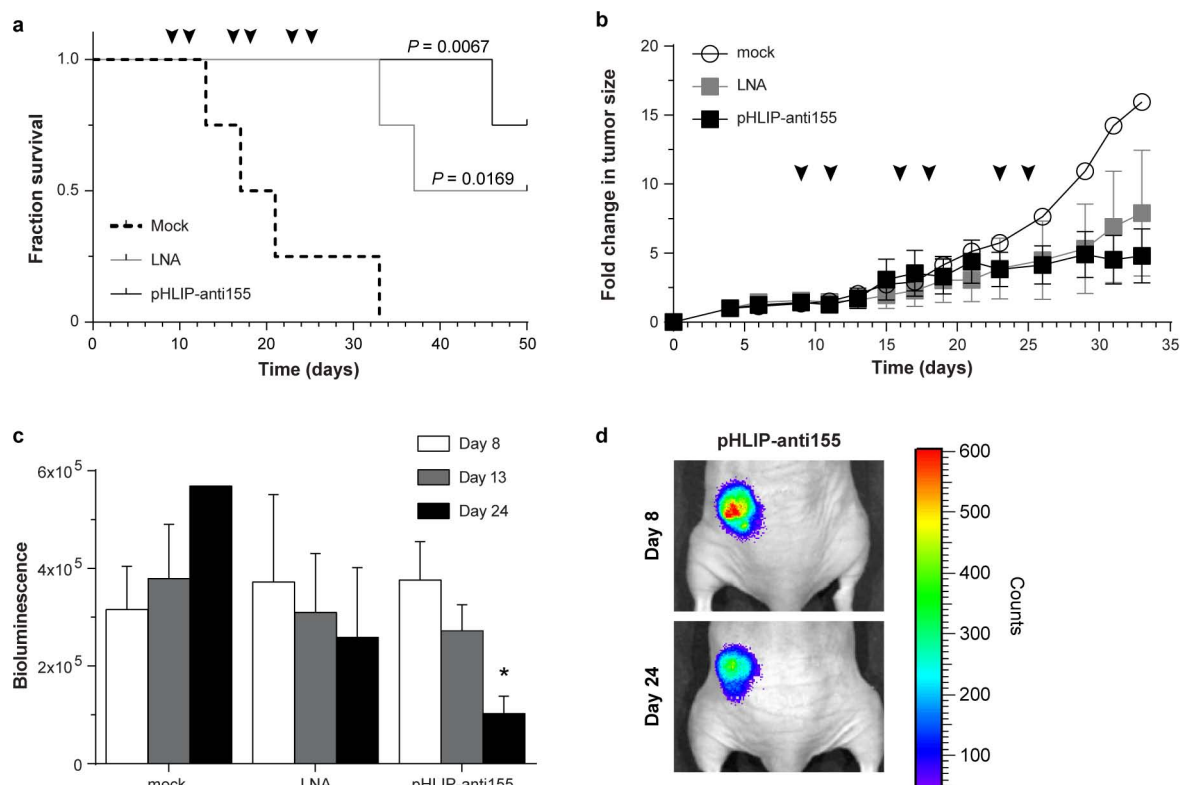
Extended Data Figure 7 | Administration of pHLIP-anti155 to mice with KB oral squamous cell carcinoma xenograft tumours. a, Intravenous injection of pHLIP-anti155 (**) and phosphorothioated LNA against miR-155 (*) significantly enhanced survival compared with mock buffer treatment; $n = 4$ for all groups; arrowheads indicate injections of 2 mg kg^{-1} (or molar equivalent for LNA). Survival cutoff criteria included tumour volume greater than 1 cm^3 or compassionate euthanasia, which was mandated for three mock-treated mice with ulcerated tumours. **b,** Fold change in tumour size in response to treatment; measurements were plotted until the mock negative control group was euthanized. **c,** Tumour bioluminescence in response to treatment; day 8 represents luciferase activity before first injection. **d,** Representative images of tumour bioluminescence. Data are shown as mean ± s.e.m.; statistical analysis performed with (**a**) Mantel-Cox analysis or (**c**) two-tailed Student's t-test, $*P < 0.05$; $**P < 0.01$.
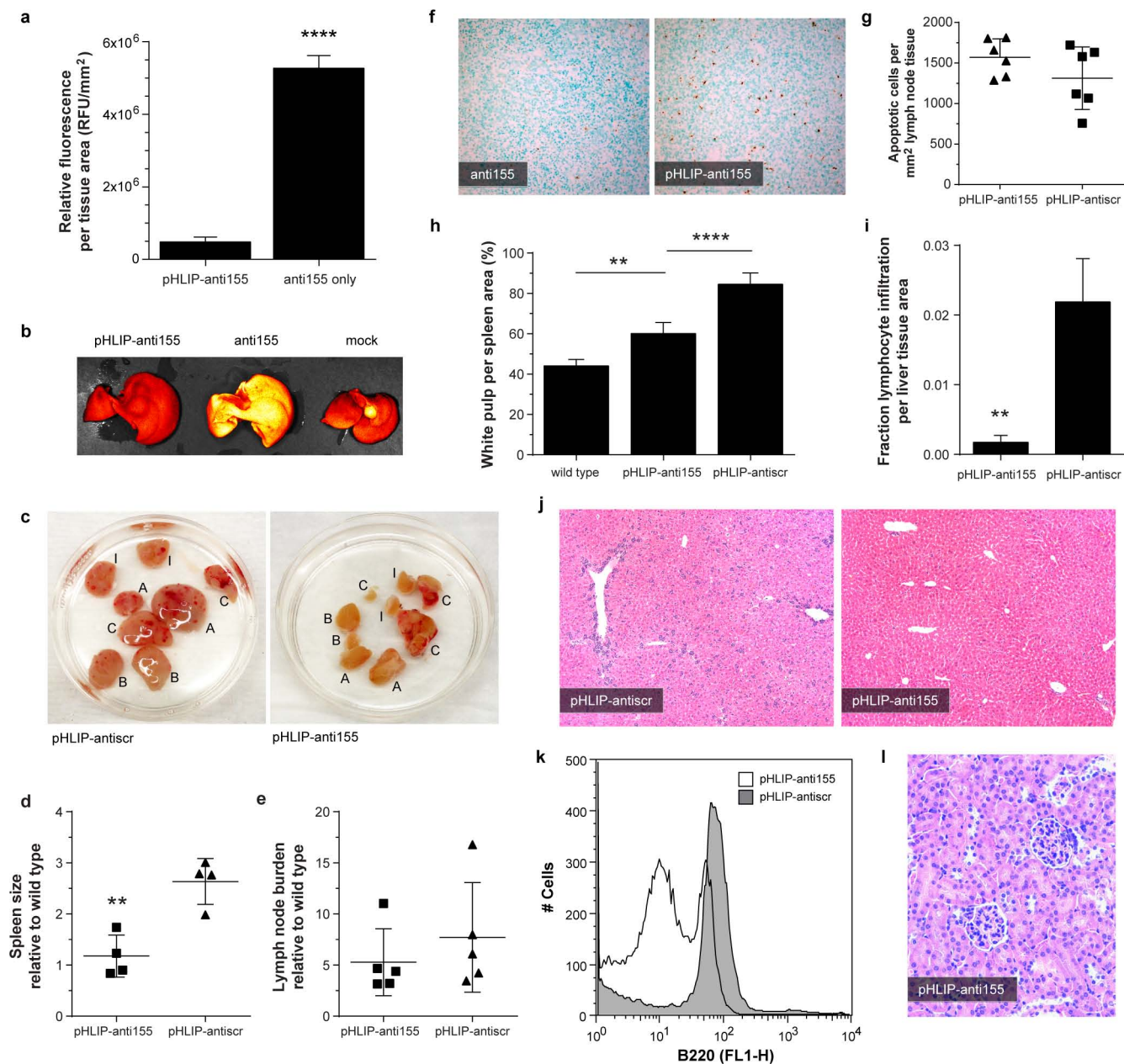
Extended Data Figure 8 | Administration of pHLIP-anti155 to *mir-155^{LSLtTA}* mice with lymphoma. **a**, Quantification of liver distribution of TAMRA-labelled PNA delivered with and without conjugation to pHLIP; ImageJ was used to measure fluorescence from five confocal sections per mouse liver; $n = 3$ mice per group. **b**, Visualization of whole liver fluorescence after antimiR administration; pHLIP-anti155 liver fluorescence is similar to the autofluorescence seen in the mock group. **c**, Lymph-node tumour burden (A, axillary; B, brachial; C, cervical; I, inguinal lymph nodes); in these specific images taken from diseased littermates, pHLIP-antiscr-treated mice had a more than threefold larger aggregate lymph node mass (3.179 g) than pHLIP-anti155-treated mice (1.006 g). **d, e**, Size of harvested (**d**) spleens ($n = 4$) and (**e**) lymph nodes (axillary, brachial, cervical, and inguinal; $n = 5$) with respect to wild type; $n < 6$ (that is, total number of treated mice) owing to size data not collected. **f, g**, TUNEL analysis of treated cervical lymph nodes of *mir-155^{LSLtTA}* mice ($n = 6$). **h**, Percentage of white pulp in treated spleens; $n = 6$. **i**, Measurement of lymphocyte infiltration into liver; $n = 6$. **j**, Low-magnification H&E images of livers from Fig. 4d. **k**, Flow cytometry analysis of B220-positive cells comprising the spleens of treated mice; B220 is typically a marker for B cells, although varied expression is seen on some T cells, natural killer cells, and macrophages; $n = 4$. **l**, Representative H&E image of healthy kidneys from pHLIP-anti155-treated mice; $n = 6$. Data are shown as mean ± s.d. (**a, d, e, g, h**) or mean ± s.e.m. (**i**); statistical analysis performed with two-tailed Student's *t*-test; **$P < 0.01$; ****$P < 0.0001$.
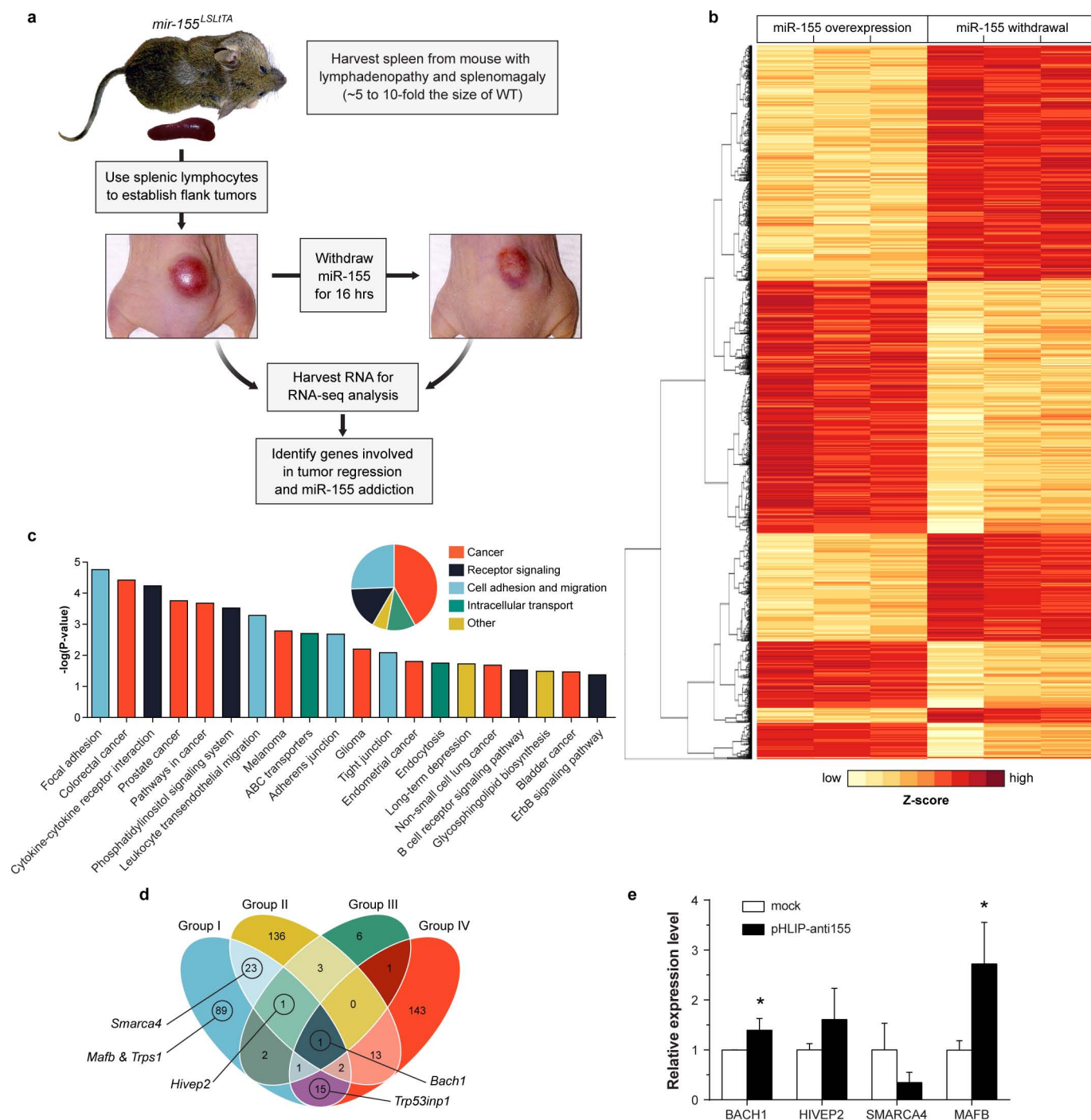
Extended Data Figure 9 | Differential gene expression analysis of miR-155 withdrawal. **a**, Experimental design for RNA-seq analysis of miR-155-addicted tumours compared with tumours undergoing miR-155 withdrawal and tumour regression. **b**, RNA-seq differential gene expression analysis of three independent tumours overexpressing miR-155 compared with three independent tumours undergoing DOX-induced miR-155 withdrawal; all differentially expressed genes with a false discovery rate less than 0.05 are shown; rows are clustered by Euclidean distance measure. **c**, KEGG pathway analysis of significantly upregulated genes after miR-155 withdrawal. **d**, Selection of potential miR-155 targets involved in tumour regression. Intersection of genes (group I) that are both predicted miR-155 targets

(Supplementary Table 2) and overexpressed after miR-155 withdrawal from *mir-155*[LSLtTA] tumours (Supplementary Table 1) with genes inferred from three separate miR-155 target analyses. Group II: the study in ref. 36 used RNA-seq to compare Mutu I B cells that overexpress miR-155 with cells transformed with a control vector[36]. Group III: ref. 25 identified shared targets between miR-155 and a viral orthologue, miR-K12-11. Group IV: the study in ref. 37 used HITS-CLIP to identify miR-155 targets without perfect seed matches in T cells. **e**, qPCR determination of gene expression levels in Toledo cells treated for 48 h with 500 nM pHLIP-anti155 at pH 6.2; data are shown as mean ± s.d.; $n = 3$; statistical analysis performed with two-tailed Student's *t*-test, *$P < 0.05$.
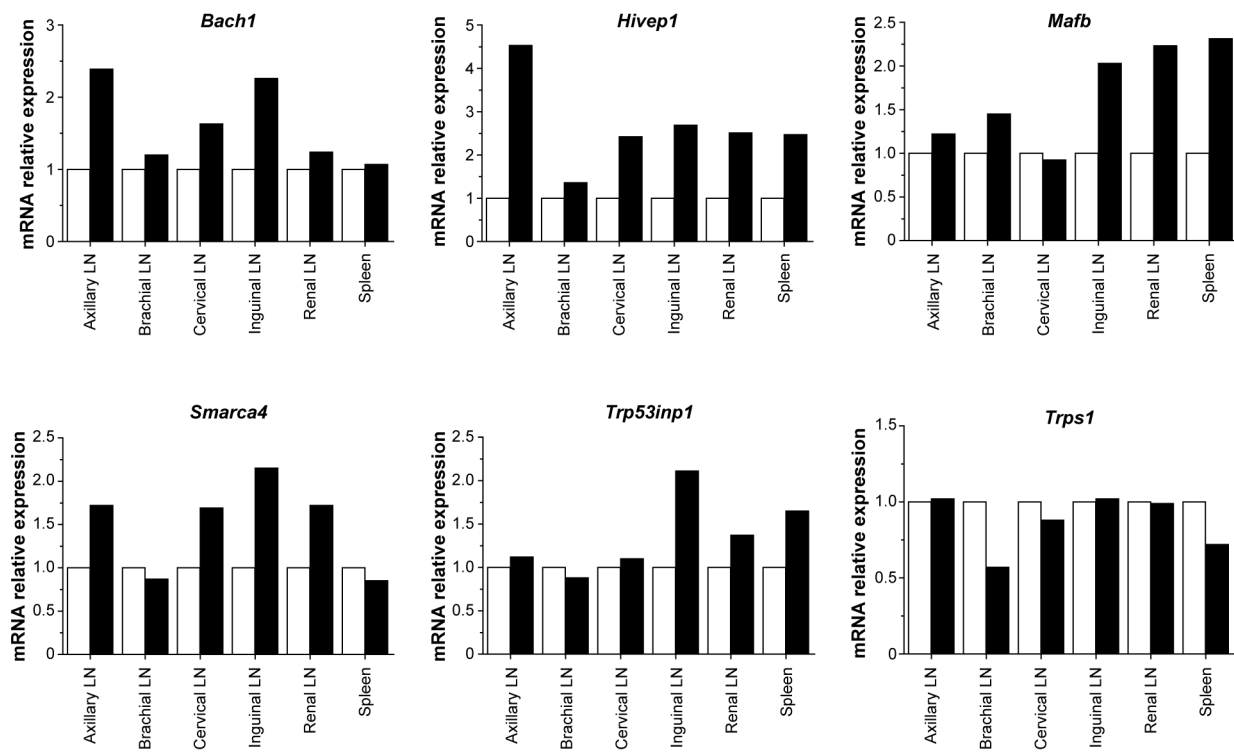
Extended Data Figure 10 | Expression levels of putative targets in response to miR-155 silencing in *mir-155^LSLtTA* mice. qPCR validation of potential miR-155 targets involved in tumour regression using *mir-155^LSLtTA* mice with conspicuous lymphadenopathy treated with (black bars) DOX for 16 h compared with (white bars) untreated mice with lymphadenopathy; all samples are normalized to β-actin; $n = 3$. Genes were selected on the basis of criteria described in Supplementary Table 3. As shown in Fig. 4f, both *Bach1* and *Mafb* have utility as biomarkers for miR-155 withdrawal-induced tumour regression.

LETTER

Optogenetic control of organelle transport and positioning

Petra van Bergeijk[1]*, Max Adrian[1]*, Casper C. Hoogenraad[1] & Lukas C. Kapitein[1]

Proper positioning of organelles by cytoskeleton-based motor proteins underlies cellular events such as signalling, polarization and growth[1–8]. For many organelles, however, the precise connection between position and function has remained unclear, because strategies to control intracellular organelle positioning with spatiotemporal precision are lacking. Here we establish optical control of intracellular transport by using light-sensitive heterodimerization to recruit specific cytoskeletal motor proteins (kinesin, dynein or myosin) to selected cargoes. We demonstrate that the motility of peroxisomes, recycling endosomes and mitochondria can be locally and repeatedly induced or stopped, allowing rapid organelle repositioning. We applied this approach in primary rat hippocampal neurons to test how local positioning of recycling endosomes contributes to axon outgrowth and found that dynein-driven removal of endosomes from axonal growth cones reversibly suppressed axon growth, whereas kinesin-driven endosome enrichment enhanced growth. Our strategy for optogenetic control of organelle positioning will be widely applicable to explore site-specific organelle functions in different model systems.

Eukaryotic cells use cytoskeletal motor proteins to control the transport and positioning of proteins, RNAs and organelles[1]. In neurons, mitochondria positioning contributes to synapse functioning and axon branching[2,7,8], whereas positioning of Golgi outposts is thought to control dendrite development[6]. Likewise, specific positioning of endosomes has been proposed to contribute to polarization and local outgrowth, either through selective delivery of building blocks or through localized signalling[5,9–12]. In many cases, however, directly resolving the role of specific organelle positioning has remained challenging. Disruption of cytoskeletal elements and inhibition of motor proteins or adaptor molecules have been frequently used to alter organelle positioning, but these approaches often lack target selectivity as well as spatial specificity. Therefore, a tool to modulate locally the distribution of specific organelles with spatiotemporal accuracy is required.

Using light-induced heterodimerization to recruit specific motors to selected cargoes might enable spatiotemporal control of intracellular transport, but whether such light-induced interactions can withstand motor-induced forces has remained unclear[13,14]. To test this, we first used light-induced binding to couple microtubule-based motors to peroxisomes in monkey COS-7 cells, because these vesicular organelles are largely immobile in the perinuclear region and any movement induced by light-targeted motor proteins could easily be observed[15]. Peroxisomes were labelled using PEX–LOV, a fusion between the peroxisomal targeting signal of PEX3 and a photosensitive LOV domain from *Avena sativa* phototropin 1, which cages a small peptide that binds the engineered PDZ domain ePDZb1 after exposure to blue light[14] (Fig. 1a, b). In addition, ePDZb1 was fused to the plus-end-directed kinesin-3 KIF1A to create KIF–PDZ. After co-expression of these two constructs and illumination with blue light, we observed the rapid redistribution of peroxisomes from the centre to the periphery of the cell where most microtubule plus-ends are located (Fig. 1c, d). Similarly, light-induced recruitment of minus-end-directed dynein using the amino terminus of

BICD2 (BICDN) fused to ePDZb1 (BICDN–PDZ) triggered the accumulation of peroxisomes at the centre of the cells (Extended Data Fig. 1a–c and Supplementary Video 1). Importantly, peroxisome redistribution did not alter the spatial organization of mitochondria, the endoplasmic reticulum, or the actin and microtubule cytoskeleton (Extended Data Fig. 2a, b).

To quantify peroxisome motility, we first used image correlation analysis to measure the overall frame-to-frame similarity before and after exposure to blue light[16]. In the absence of transport, two subsequent images are largely identical and the correlation index will be close to 1, whereas a value of 0 indicates that all organelles have moved to previously unoccupied positions. After light-induced recruitment of KIF1A, the correlation index rapidly decreased from 0.97 ± 0.01 (mean $\pm$ s.e.m.)
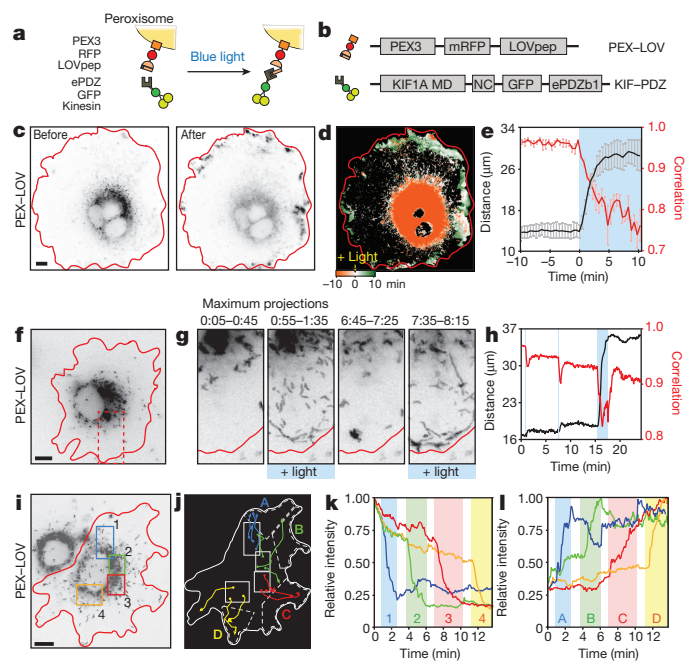
Figure 1 | Local and reversible activation of microtubule-based transport with light. **a, b,** Assay and constructs. MD, motor domain; NC, neck coil. **c,** Peroxisome distribution before and after light-induced recruitment of KIF–PDZ. **d,** Colour-coded overlay of time series. **e,** Displacement (black, expressed in $R_{90\%}$) and correlation (frame-to-frame similarity from 0 to 1, red) versus time ($n = 6$ cells, mean $\pm$ s.e.m.). Blue marks illumination. **f, g,** Reversible activation using pulsed light. **g,** Maximum intensity projections during periods of 40 s. See Supplementary Video 2. Time shown in minutes:seconds. **h,** Displacement (black, $R_{90\%}$) and correlation (red) versus time. **i–l,** Local activation using sequential illumination of four regions (**i**), resulting in outward targeting to adjacent regions (**j**, showing example trajectories), quantified using normalized fluorescence intensity (**k, l,** coloured boxes mark blue-light illumination). See Supplementary Video 3. Scale bars, 10 μm.

[1]Cell Biology, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands.
*These authors contributed equally to this work.

to 0.76 ± 0.04, reflecting the induction of continuous peroxisome motility (Fig. 1e). By contrast, dynein recruitment eventually increased the correlation index, because most peroxisomes accumulated at the same position in the centre of the cell (Extended Data Fig. 1c, d). To quantify this overall peroxisome repositioning, we calculated for each time point the radius of the circle required to enclose 90% of the fluorescence intensity of the peroxisomes, $R_{90\%}$, and found a large increase from 14 ± 2 μm to 29 ± 3 μm on recruitment of KIF1A (Fig. 1e). By contrast, $R_{90\%}$ decreased from 15.4 ± 0.3 μm to 12.8 ± 0.6 μm on recruitment of dynein (Extended Data Fig. 1d). Thus, rapid organelle redistribution can be induced by using light to recruit microtubule motors.

To achieve spatiotemporal control, recruitment of motors should be both reversible and locally restricted. To test the reversibility of motility induction, we exposed cells expressing KIF–PDZ and PEX–LOV to three consecutive periods of blue light, interspersed with ~7 min without blue-light exposure. Whereas peroxisomes moved outwards during blue-light illumination, movement was arrested within seconds without blue light (Fig. 1f, g and Supplementary Video 2). By contrast, $R_{90\%}$ remained stable without stimulation (Fig. 1h), indicating that peroxisomes do not spontaneously return to their original position after motor dissociation (Extended Data Fig. 3a). To test whether transport could be induced locally, we sequentially illuminated four different regions within a cell (Fig. 1i, j). Peroxisomes in the activated region rapidly redistributed to non-exposed areas, whereas non-exposed peroxisomes remained stationary (Fig. 1j and Supplementary Video 3). The fluorescence intensity in the illuminated boxes 1–4 decreased by 60–75%, coinciding with a 180–280% increase in the adjacent peripheral boxes A–D (Fig. 1j–l). These results demonstrate that transport of intracellular cargo can be induced with spatiotemporal precision.

We have previously shown that myosin-V can oppose kinesin-driven transport in actin-dense regions[16], suggesting that light-induced recruitment of myosin-V can be used to anchor organelles at specific sites. To test this, myosin-Vb was recruited to peroxisomes preloaded with the kinesin-2 KIF17 (refs 15, 16) (Fig. 2a, b). Whereas the attached kinesin motor ensured continuous motility of many peroxisomes near the

cell periphery (Fig. 2c), this motility was arrested after recruitment of myosin-Vb, resulting in a 30% increase of the correlation index (Fig. 2d). Local illumination increased the correlation index to similar levels, but only in the exposed region (Fig. 2c, d). Moreover, individual peroxisome trajectories showed on average four times smaller frame-to-frame displacements during illumination compared to before and after stimulation (Fig. 2e, f). These data demonstrate that organelle motility can be stalled with spatiotemporal precision through light-induced recruitment of myosin-Vb.

We next used RAB11-positive recycling endosomes to test our method on organelles whose proper physiological functioning depends on selective transport and positioning. Kinesin- and dynein-based redistribution and myosin-Vb-based anchoring of recycling endosomes could be transiently induced with light (Extended Data Figs 2–4, see also Supplementary Information, Extended Data Figs 5 and 6 and Supplementary Videos 4 and 5), demonstrating that the movement of intrinsically dynamic cargoes can be temporarily amplified or overruled by coupling these cargoes to a specific motor using light. Notably, whereas peroxisomes remained largely immobile at their new location after light-dependent repositioning, the original distribution of recycling endosomes was quickly restored after the light-induced kinesin recruitment was stopped (Extended Data Fig. 3).

To test our approach in a more complex and delicate model system, we switched to primary cultures of rat hippocampal neurons. Their polarized morphology and specialized cytoskeletal organization in different compartments, such as axons, dendrites and dendritic spines, should allow transporting cargoes into and out of these compartments by recruiting the appropriate motor proteins. We first examined whether light-induced recruitment of myosin-Vb was sufficient to drive transport into dendritic spines, as proposed previously[17–20]. Indeed, in cells co-expressing PEX–LOV and a fusion of myosin-Vb with ePDZb1 (MYO–PDZ), 62 ± 3% of the illuminated spines were targeted with peroxisomes compared to 1 ± 1% of spines in non-illuminated dendrites (Fig. 2g–j and Supplementary Video 6). After the illumination period, the number of peroxisome entries decreased with a half-time of ~36 s (Fig. 2j).
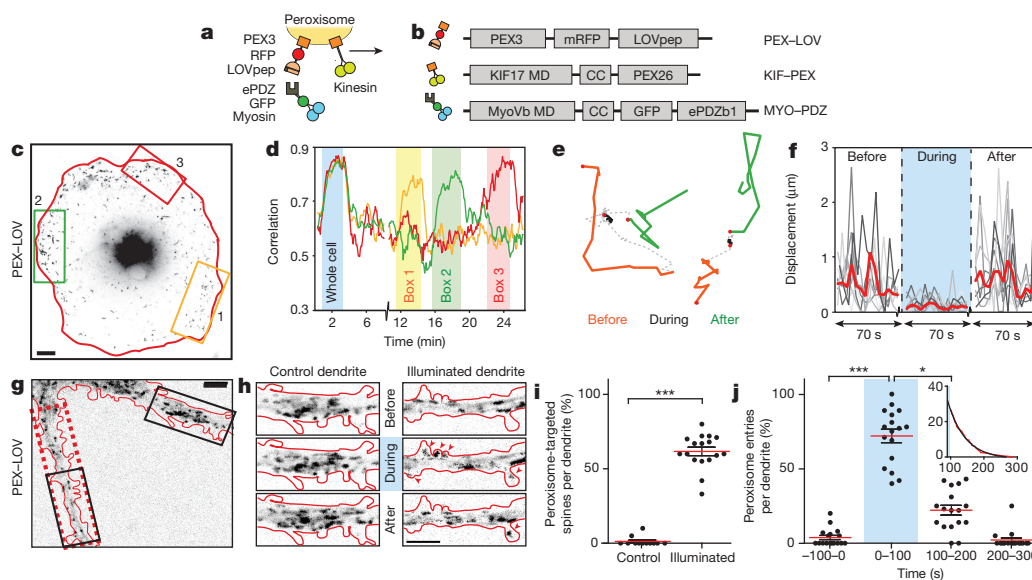
Figure 2 | Light-induced myosin-Vb recruitment anchors organelles or targets them into dendritic spines. a, b, Assay and constructs. CC, coiled coil. **c**, Peroxisome distribution in cell expressing PEX–LOV, KIF–PEX and MYO–PDZ. **d**, Correlation time trace for areas shown in **c**. **e**, Peroxisome trajectories with 70-s episodes before, during and after myosin-Vb recruitment. **f**, Frame-to-frame displacements of peroxisomes (5 s interval). Red denotes the average of nine individual peroxisome trajectories. **g**, Peroxisome distribution in primary hippocampal neuron expressing PEX–LOV and MYO–PDZ. Dashed red rectangle was illuminated. **h**, Thirty-second maximum

projections of regions from **g**. Arrowheads mark peroxisomes in spines. See Supplementary Video 6. **i**, Spine targeting in control ($n = 12$) and illuminated ($n = 17$) dendrites, in three independent experiments. Mean ± s.e.m., ***$P < 0.0001$, Mann–Whitney test. **j**, Spine entries over time. Mean ± s.e.m., *$P < 0.05$, ***$P < 0.0001$, Kruskal–Wallis analysis of variance (ANOVA), Dunn's post-hoc test, $n = 17$ dendrites. Inset: entry probability after illumination (red) fitted with exponential decay $\exp(t\lambda^{-1})$ (black, $\lambda = 36.36$ s). Scale bars, 5 μm (**g, h**) and 10 μm (**c**).

Similarly, RAB11 recycling endosomes could be enriched in specific spines by local illumination (Extended Data Fig. 4o–q), demonstrating that light-controlled transport can be used to manipulate individual dendritic spines.

RAB11 vesicles have been implicated in the control of axon growth, but their local role in the growth cone could not be assessed previously[11,12,21]. We therefore used local light-induced recruitment of motor proteins to RAB11 recycling endosomes to test how local dynein-driven removal or kinesin-driven addition of endosomes affects growth cone dynamics (Fig. 3a, d). Importantly, neither illumination nor addition of the heterodimerizer rapalog (used to link the LOV domain to the N terminus of RAB11) altered growth cone structure or behaviour in cells expressing FRB–LOV and PDZ only (Extended Data Figs 5 and 7). Likewise, in control neurons expressing FKBP–RAB11 together with BICDN–PDZ or KIF–PDZ, but lacking the FRB–LOV protein, exposure to blue light did not affect the rapid filopodial and lamellipodial dynamics or the overall growth of most growth cones. (Fig. 3b, e). When dynein was coupled to RAB11, a clear decrease in growth cone dynamics and growth was observed (Fig. 3c, g, Extended Data Fig. 8 and Supplementary Video 7). By contrast, coupling of kinesin resulted in rapid axon extension in $39 \pm 7\%$ of the growth cones (Fig. 3f, h and Supplementary Video 7). Importantly, when growth cones were not completely collapsed upon dynein-dependent RAB11 depletion, this depletion and the reduced growth cone dynamics could be reversed when cells were no longer exposed to blue light (Fig. 3i, j). These data demonstrate that growth cone dynamics and axon growth d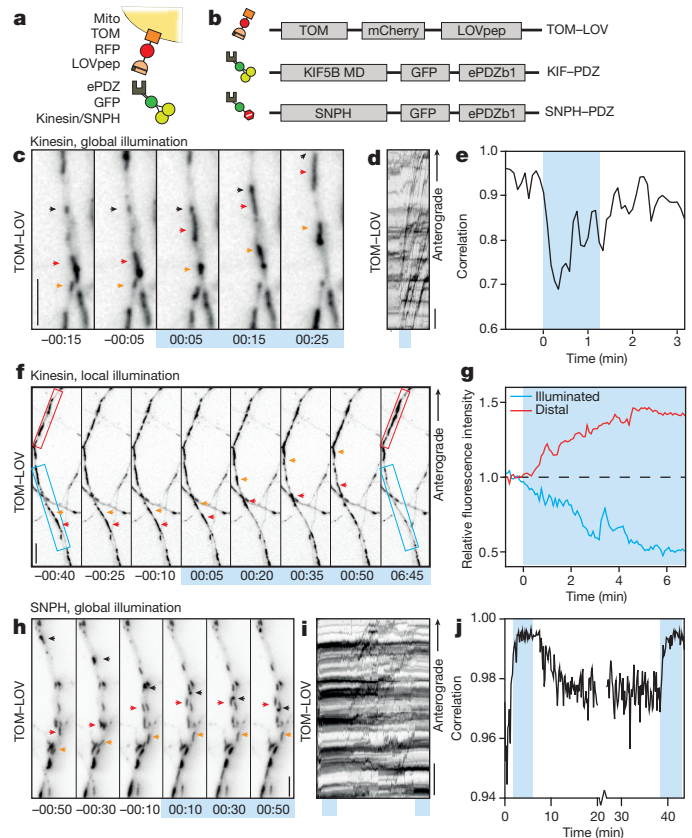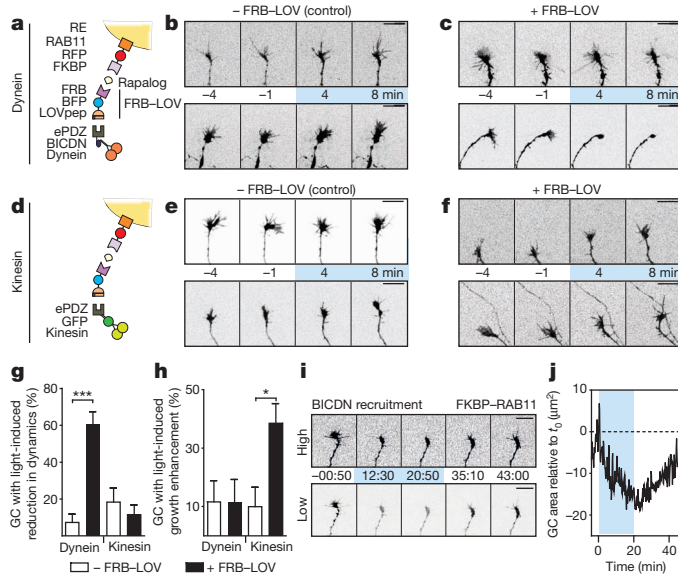irectly depend on RAB11 vesicle functioning near the growth cone, rather than on general RAB11 functions elsewhere in the cell.

Recently, the controlled anchoring and mobilization of mitochondria have emerged as key regulatory events in neurons[8,22–24]. Mitochondrial positioning depends on both motor-dependent transport and controlled immobilization by specific docking factors, but the molecular and mechanical interplay between motors and docking factors has remained

unclear[22,24]. For example, syntaphilin (SNPH) has been proposed to induce anchoring by crosslinking mitochondria to microtubules and through a direct inactivating interaction with kinesin[25,26], but whether remobilization requires the regulated release of both interactions is not known[23]. To test whether recruitment of more motors can overcome anchoring, we used light to recruit KIF–PDZ to axonal mitochondria labelled with TOM–LOV (Fig. 4a, b) and found that this was sufficient to mobilize most mitochondria in the illuminated region (Fig. 4c–g and Supplementary Video 8). Conversely, light-induced recruitment of the N-terminal part of SNPH was sufficient to acutely anchor motile mitochondria, independent of their directionality (Fig. 4h–j and Supplementary Video 9). These results demonstrate that regulation of mitochondrial motility and anchoring does not require all-or-nothing switching between the activation and inactivation of specific motors, but instead depends on the balance of forces between active motors and passive anchors.

We have established optically-controlled intracellular transport by using light-sensitive heterodimerization to recruit specific cytoskeletal motor proteins to selected cargoes. Our ability to control organelle positioning complements recent work that established optogenetic control over nucleocytoplasmic distribution of proteins[27]. We anticipate that this approach will be widely applicable to study how organelle positioning controls cellular functioning, as demonstrated here for the role

Figure 3 | Motor-based redistribution of recycling endosomes modulates axon outgrowth. a, d, Assay and constructs. Rapalog targets FRB–LOV to the RAB11 N terminus. **b, c, e, f,** Growth cone dynamics of neurons expressing FKBP–RAB11 and BICDN–PDZ (**b, c**) or KIF–PDZ (**e, f**) without (**b, e**) or with (**c, f**) FRB–LOV. See Supplementary Video 7. **g, h,** Light-induced reduction of growth cone (GC) dynamics (**g**) or light-induced growth enhancement (**h**). Dynein/kinesin: −FRB–LOV: $n = 21/19$ axons, +FRB–LOV: $n = 25/35$ axons, in $n = 5/5$ independent experiments. Mean ± s.e.m., $*P < 0.05$, $***P < 0.0001$, unpaired two-tailed t-test. **i,** Same growth cone in low and high contrast illustrating reversibility of reduced FKBP–RAB11 targeting and growth cone dynamics. **j,** Area over time for the example shown in **i**. Scale bars, 10 µm.

Figure 4 | Altering mitochondrial dynamics through recruitment of motors and anchors. a, b, Assay and constructs. **c,** Axonal mitochondria before and during KIF–PDZ recruitment. Arrowheads track individual mitochondria. See Supplementary Video 8. **d,** Kymograph for axon shown in **c**, representative for $n = 6$ axons. Blue box marks activation (1 min:15 s). **e,** Correlation over time for region shown in **c**. **f,** Axonal mitochondria before and during local illumination (blue box). See Supplementary Video 8. **g,** Relative fluorescence intensity versus time in the illuminated (blue box in **f**) and the adjacent, distal region (red box in **f**). **h,** Axonal mitochondria before and during SNPH recruitment. Arrowheads track individual mitochondria. See Supplementary Video 9. **i,** Kymograph for axon shown in **h**, representative for $n = 5$ axons. Blue boxes mark activation (4 min:50 s). **j,** Correlation versus time for region in Supplementary Video 9. Scale bars, 5 µm (**c, f, h**) and 10 µm (**d, i**).

of recycling endosomes in growth cone dynamics. In addition, it could be used to control cellular processes such as polarization, signalling and outgrowth by depleting or accumulating cargo at specific sites. For example, increased axonal targeting of certain cargoes might promote axon regeneration after injury and provide novel insights into the mechanisms contributing to regeneration failure or success, both in culture and in different animal models[28].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112**, 467–480 (2003).
2. Sheng, Z. H. & Cai, Q. Mitochondrial transport in neurons: impact on synaptic homeostasis and neurodegeneration. *Nature Rev. Neurosci.* **13**, 77–93 (2012).
3. Korolchuk, V. I. *et al.* Lysosomal positioning coordinates cellular nutrient responses. *Nature Cell Biol.* **13**, 453–460 (2011).
4. Yadav, S. & Linstedt, A. D. Golgi positioning. *Cold Spring Harb. Perspect. Biol.* http://dx.doi.org/10.1101/cshperspect.a005322 (2011).
5. Sadowski, L., Pilecka, I. & Miaczynska, M. Signaling from endosomes: location makes a difference. *Exp. Cell Res.* **315**, 1601–1609 (2009).
6. Ori-McKenney, K. M., Jan, L. Y. & Jan, Y. N. Golgi outposts shape dendrite morphology by functioning as sites of acentrosomal microtubule nucleation in neurons. *Neuron* **76**, 921–930 (2012).
7. Spillane, M., Ketschek, A., Merianda, T. T., Twiss, J. L. & Gallo, G. Mitochondria coordinate sites of axon branching through localized intra-axonal protein synthesis. *Cell Rep.* **5**, 1564–1575 (2013).
8. Courchet, J. *et al.* Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. *Cell* **153**, 1510–1525 (2013).
9. Golachowska, M. R., Hoekstra, D. & van IJzendoorn, S. C. Recycling endosomes in apical plasma membrane domain formation and epithelial cell polarity. *Trends Cell Biol.* **20**, 618–626 (2010).
10. Higuchi, Y., Ashwin, P., Roger, Y. & Steinberg, G. Early endosome motility spatially organizes polysome distribution. *J. Cell Biol.* **204**, 343–357 (2014).
11. Eva, R. *et al.* ARF6 directs axon transport and traffic of integrins and regulates axon growth in adult DRG neurons. *J. Neurosci.* **32**, 10352–10364 (2012).
12. Eva, R. *et al.* Rab11 and its effector Rab coupling protein contribute to the trafficking of β1 integrins during axon growth in adult dorsal root ganglion neurons and PC12 cells. *J. Neurosci.* **30**, 11654–11669 (2010).
13. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods* **7**, 973–975 (2010).
14. Strickland, D. *et al.* TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nature Methods* **9**, 379–384 (2012).
15. Kapitein, L. C. *et al.* Probing intracellular motor protein activity using an inducible cargo trafficking assay. *Biophys. J.* **99**, 2143–2152 (2010).
16. Kapitein, L. C. *et al.* Myosin-V opposes microtubule-based cargo transport and drives directional motility on cortical actin. *Curr. Biol.* **23**, 828–834 (2013).
17. Correia, S. S. *et al.* Motor protein-dependent transport of AMPA receptors into spines during long-term potentiation. *Nature Neurosci.* **11**, 457–466 (2008).
18. Wagner, W., Brenowitz, S. D. & Hammer, J. A. III. Myosin-Va transports the endoplasmic reticulum into the dendritic spines of Purkinje neurons. *Nature Cell Biol.* **13**, 40–48 (2011).
19. Wang, Z. *et al.* Myosin Vb mobilizes recycling endosomes and AMPA receptors for postsynaptic plasticity. *Cell* **135**, 535–548 (2008).
20. Hammer, J. A. III & Wagner, W. Functions of class V myosins in neurons. *J. Biol. Chem.* **288**, 28428–28434 (2013).
21. Bhuin, T. & Roy, J. K. Rab11 is required for embryonic nervous system development in *Drosophila. Cell Tissue Res.* **335**, 349–356 (2009).
22. Encalada, S. E. & Goldstein, L. S. Biophysical challenges to axonal transport: motor-cargo deficiencies and neurodegeneration. *Ann. Rev. Biophy.* **43**, 141–169 (2014).
23. Sheng, Z. H. Mitochondrial trafficking and anchoring in neurons: new insight and implications. *J. Cell Biol.* **204**, 1087–1098 (2014).
24. Sun, T., Qiao, H., Pan, P. Y., Chen, Y. & Sheng, Z. H. Motile axonal mitochondria contribute to the variability of presynaptic strength. *Cell Rep.* **4**, 413–419 (2013).
25. Chen, Y. & Sheng, Z. H. Kinesin-1-syntaphilin coupling mediates activity-dependent regulation of axonal mitochondrial transport. *J. Cell Biol.* **202**, 351–364 (2013).
26. Kang, J. S. *et al.* Docking of axonal mitochondria by syntaphilin controls their mobility and affects short-term facilitation. *Cell* **132**, 137–148 (2008).
27. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5**, 4404 (2014).
28. Bradke, F., Fawcett, J. W. & Spira, M. E. Assembly of a new growth cone after axotomy: the precursor to axon regeneration. *Nature Rev. Neurosci.* **13**, 183–193 (2012).

METHODS

DNA constructs. The following constructs have been described: tagRFPt[29] (gift from R. Tsien), pCIBN(deltaNLS)-pmGFP and pCry2PHR-mCherryN1 (ref. 13; addgene, plasmids 26867 and 26866), mid(SS/TM)-GFP-LOVpep and ePDZb1-mCherry[14] (addgene, plasmids 34972 and 34981), TOM20-mCherry-GAI[30] (gift from T. Inoue), HA-rab11a (ref. 31), Kif1a(1–489)-GFP-FRB, Kif5b(1–807)-GFP-FRB, MyoVb(1–1090)-GFP-FRB, HA-BicD2(1–594)-FRB and Pex3(1–42)-mRFP[15], pGW2-pex26 and pGW2-Kif17-GFP-pex26 (ref. 16), GFP-SNPH[26] (gift from Z. Sheng), GFP-RCP[32] (RAB11FIP1, gift from R. Prekeris), NPY-GFP[33] and GFP-MACF18 (ref. 34).

Cloning vectors and fluorescent tags. The constructs used in this study were cloned in the mammalian expression vectors pGW1-CMV, pGW2-CMV and/or pβactin[15]. pβactin-GFP, pβactin-tagRFPt, pβactin-tagBFP and pβactin-iRFP were generated by ligating the fluorescent proteins in the SalI and SpeI site of pβactin.

Tagging motor proteins, adaptors and anchors with CIBN and ePDZb1. To generate myoVb(1–1090)-GFP-CIBN and myoVb(1–1090)-GFP-ePDZb1, amino acids 1–1090 of myosin-Vb were cloned in the AscI and EcoRI sites of pβactin-GFP, and either CIBN or ePDZb1 was inserted downstream of GFP using a PCR-based strategy. Similarly, myoVb-iRFP-CIBN was made using the pβactin-iRFP vector backbone. Kif1a(1–383)-GFP-CIBN and Kif1a(1–383)-GFP-ePDZb1 were generated by ligating amino acids 1–383 of mouse KIF1A in the AscI and SalI sites of pβactin-GFP. Subsequently, PCR-amplified CIBN or ePDZb1 was inserted downstream of GFP. Haemagglutinin (HA)-tagged HA-BicD2(1–500)-CIBN and BicD2(1–500)-ePDZb1 were cloned by inserting PCR-amplified BicD2(1–500) (refered to as BICDN in the main text) into the pβactin vector backbone. Subsequently, CIBN and ePDZb1were ligated downstream of BicDN. Kif5b(1–807)-GFP-ePDZb1 was made by inserting PCR-amplified Kif5b(1–807) into the AscI and BamHI sites of a GFP-ePDZb1 backbone. To create SNPH(45–748)-GFP-ePDZb1, PCR-amplified *SNPH* (forward primer: 5′-AGCGCTAAGCTTGCCACCATGGCCATGTCCCT GCAGGGAAG-3′ and reverse primer: 5′-GCCCTTGCTCACCATAGTCGACC CCACTACCACAGCCAGCAGATCCAC-3′) was inserted into a GFP-ePDZb1 backbone using Cold Fusion cloning (System Biosciences).

Tagging peroxisomes, RAB11 vesicles and mitochondria with LOVpep and Cry2PHR. To generate Pex3-mRFP-LOVpep (PEX–LOV), LOVpep, including a 9-amino-acid linker (GGSGGSGGS), was ligated in the AscI and SalI sites of pGW1-pex3-mRFP. To make TOM20-mCherry-LOVpep, Pex3-mRFP was replaced by TOM20(1–34)-mCherry using the HindIII and AscI sites. To create Cry2PHR-tagRFPt-Rab11 and FKBP-tagRFPt-Rab11, Rab11a was introduced in the SpeI and NotI sites of pβactin-tagRFPt. Subsequently, PCR-amplified FKBP or Cry2PHR was ligated upstream of tagRFPt. FRB-tagBFP-LOVpep was made by inserting LOVpep, including a 9-amino-acid linker, in the SpeI and NotI sites of pβactin-tagBFP. Subsequently, PCR -amplified FRB was cloned upstream of tagBFP.

Other constructs. pJPA5-TfR-GFP (a gift from G. Banker) was cloned into a β-actin vector. Membrane targeting of GFP was achieved by inserting the 40 most N-terminal residues of the MARCKS protein with an additional palmitoylation site at residue 3 (ref. 35) into GW2-tagRFPt. To generate mRFP-actin, human cytoplasmic β-actin was cloned from pEGFP-actin (Clontech) in the β-actin-mRFP vector.

Cell cultures and transfection. COS-7 cells were cultured in DMEM/Ham's F10 (1:1) medium containing 10% FCS and penicillin/streptomycin. Then, 2–4 days before transfection, cells were plated on 24-mm diameter coverslips. Cells were transfected with Fugene6 transfection reagent (Roche) according to the manufacturer's protocol and imaged one day after transfection.

Primary hippocampal cultures were prepared from embryonic day 18 (E18) rat brains[36]. Cells were plated on coverslips coated with poly-L-lysine (30 mg ml^{-1}) and laminin (2 mg ml^{-1}) at a density of 75,000 per well. Hippocampal cultures were grown in Neurobasal medium (NB) supplemented with B27, 0.5 mM glutamine, 12.5 mM glutamate, and penicillin plus streptomycin. Hippocampal neurons were transfected 48 h before imaging with lipofectamine 2000 (Invitrogen). DNA (3.6 μg per well) was mixed with 6.6 μl lipofectamine 2000 in 400 ml NB, incubated for 30 min, and then added to the neurons in NB supplemented with 0.5 mM glutamine at 37 °C in 5% CO$_2$ for 60 min. Next, neurons were washed with NB and transferred to the original medium at 37 °C in 5% CO$_2$ for 2 days. Transport assays targeting dendritic spines were imaged at day-*in-vitro* (DIV) 20–22 and growth cone or mitochondria assays were imaged at DIV 3–7.

Live-cell image acquisition. Time-lapse live-cell imaging of COS-7 cells and hippocampal neurons was performed on a Nikon Eclipse TE2000E (Nikon) equipped with an incubation chamber (Tokai Hit; INUG2-ZILCS-H2) mounted on a motorized stage (Prior)[16]. Coverslips (24 mm) were mounted in metal rings, immersed in 0.6 ml Ringer's solution (10 mM HEPES, 155 mM NaCl, 5 mM KCl, 1 mM CaCl$_2$, 1 mM MgCl$_2$, 2 mM NaH$_2$PO$_4$ and 10 mM glucose, pH 7.4) or full medium (RAB11 imaging in COS-7 cells) or conditioned medium (Neuron imaging), and maintained at 37 °C and 5% CO$_2$. Cells were imaged every 5, 10 or 30 s for 5–50 min using a 40× objective (Plan Fluor, numerical aperture (NA) 1.3, Nikon) and a Coolsnap HQ2 CCD camera (Photometrics). Dense-core vesicles were imaged using a 100×

objective (Apo TIRF, 1.49 NA, Nikon) on a Evolve 512 EMCCD camera (Photometrics). A mercury lamp (Osram) and filter wheel containing ET-GFP (49002), ET-dsRed (49005), ET-mCherry (49008) and ET-GFPmCherry (59022) emission filters (all Chroma) were used for excitation and for global activation. For global activation, the GFP excitation filter was used to illuminate the sample for 100–150 ms with every image acquisition during the periods of blue-light exposure. In most experiments, the activation intensity was around 10 W cm^{-2} and the total irradiance was about 30 times higher than the minimum irradiance required for full activation (see Extended Data Fig. 4h). These settings allowed us to monitor the dynamics of GFP-labelled proteins or growth cones during activation.

For local illumination of specific areas using a 488-nm laser line, a FRAP scanning head was used (FRAP L5 D-CURIE, Curie Institute). Compared to standard FRAP experiments the laser was used at much lower intensities.

Live-cell spinning disk confocal microscopy of growth cones and spines of hippocampal neurons was performed on a Nikon Eclipse-Ti (Nikon) microscope with a Plan Apo VC, 60×, 1.40 NA oil objective (Nikon). The microscope is equipped with a motorized stage (ASI; PZ-2000), a Perfect Focus System (Nikon), an incubation chamber (Tokai Hit; INUG2-ZILCS-H2) and uses MetaMorph 7.7.11 software (Molecular Devices) to control the camera and all motorized parts. Confocal excitation and detection is achieved using 100 mW Vortran Stradus 405 nm, 100 mW Cobolt Calypso 491 nm and 100 mW Cobolt Jive 561nm lasers and a Yokogawa spinning disk confocal scanning unit (CSU-X1-A1N-E; Roper Scientific) equipped with a triple-band dichroic mirror (z405/488/568trans-pc; Chroma) and a filter wheel (CSUX1-FW-06P-01; Roper Scientific) containing 4′,6-diamidino-2-phenylindole (DAPI; ET-DAPI (49000)), GFP (ET-GFP (49002)) and mCherry (ET-mCherry (49008)) emission filter (all Chroma). Confocal images were acquired with a QuantEM:512 SC EMCCD camera (Photometrics) at a final magnification of 110 nm per pixel, including the additional 2.5× magnification introduced by an additional lens mounted between scanning unit and camera (VM Lens C-2.5X; Nikon). Local activation of photo-heterodimerization was achieved with an ILas FRAP system (Roper Scientific France/ PICT-IBiSA, Institut Curie) and 491 nm laser line at low power. To couple FRB-LOV to FKBP-RAB11, rapalog (AP21967, ARIAD) was dissolved to 0.1 mM in ethanol. Then 20 min before imaging, 0.2 ml of culture medium with rapalog (400 nM) was added to establish a final rapalog concentration of 100 nM.

Image processing and analysis. Images of live cells were processed and analysed using MetaMorph (Molecular Devices), LabVIEW (National Instruments) software and ImageJ (NIH). If not followed by a quantification in a subsequent panel, representative images are representative of 60–90% of the cells studied in the same conditions, with at least five responding cells per condition (except for Extended Data Fig. 4f with three responding cells, because we used the system in Extended Data Fig. 4i, j for all subsequent experiments). The exact organelle distributions and dynamics mostly depended on the levels of protein overexpression, which could not be examined before the experiment without triggering heterodimerization. For example, if the motor were poorly expressed, less redistribution was observed. This was most apparent in experiments where three or more constructs were co-expressed, some of which without fluorescent marker that could be used to confirm expression of the motor.

Quantification of redistribution dynamics. Before analysis, cells were masked to exclude contributions from neighbouring cells to the analysis. For the colour-coded redistribution plots, all images of a time-lapse recording were thresholded at ~5–20 times the standard deviation of the background above the background to yield binary images that were subsequently overlaid non-transparently starting with the final frame (first frame on top) in Fig. 1d, and starting with the first frame (last frame on top) in Fig. 3c. Each frame was coloured using a time-coded gradient that ran from orange to white before and from white to green after blue-light illumination. To quantify the radial redistribution of peroxisomes upon recruitment of (additional) motor proteins, the radius required to include 90% of the total intensity of the cell, $R_{90\%}(t)$, was calculated for each frame as described previously[15].

To quantify changes in the dynamics of peroxisomes or RAB11 vesicles upon recruitment of (additional) motor proteins, we calculated the time-dependent frame-to-frame correlation index $c_\tau(t)$[16] by first calculating the integrated intensity of the image obtained by multiplying the frames acquired at t and $t + \Delta t$, that is, $\sum_{x=1}^{X}\sum_{y=1}^{Y} i(x,y,t)i(x,y,t+\Delta t)$, in which $i(x, y, t)$ is the intensity at pixel (x, y) of the frame acquired at time t. These values can then be normalized using either the integrated intensity of the image acquired at t or $t + \Delta t$, that is, $\sum_{x=1}^{X}\sum_{y=1}^{Y} i^2(x,y,t)$ or $\sum_{x=1}^{X}\sum_{y=1}^{Y} i^2(x,y,t+\Delta t)$. For our analyis, we used the average result of these two normalizations as our readout, that is,

$$c_\tau(t) = \frac{1}{2}\left(\frac{\sum_{x=1}^{X}\sum_{y=1}^{Y} i(x,y,t)i(x,y,t+\tau)}{\sum_{x=1}^{X}\sum_{y=1}^{Y} i^2(x,y,t)} + \frac{\sum_{x=1}^{X}\sum_{y=1}^{Y} i(x,y,t)i(x,y,t+\tau)}{\sum_{x=1}^{X}\sum_{y=1}^{Y} i^2(x,y,t+\tau)}\right).$$

$c_\tau(t)$ will be 1 if the particles are completely anchored and their positions unchanged after a time τ, whereas $c_\tau(t)$ will be 0 if all particles moved to previously unoccupied locations. In practice, $c_\tau(t)$ will remain finite even in very dynamic samples, because a subset of particles will move to locations that were occupied by different particles in the first image. In all our analyses, we used frame-to-frame differences. For analysing the correlation index in small regions (Figs 2d and Extended Data Fig. 4l), measurements were averaged over six adjacent time points.

To determine local changes in fluorescence intensities over time (Fig. 1i–l), the mean grey value of the first-frame-subtracted recording (Fig. 1l), or last-frame-subtracted recording (Fig. 1k) was measured, and the maximum was set to 1. Individual peroxisomes or RAB11 vesicle trajectories were obtained using the MTrackJ plugin in ImageJ[37].

Analysis of spine entries. Peroxisomes and RAB11 vesicles were imaged at 1-s intervals, preferably in two dendrites of the same neuron, of which one was illuminated with pulses of 491 nm light directly before the frames indicated. Spine entries during periods of 100 s before, 100 s during and 200 s after illumination were detected manually using the Cell Counter plugin in ImageJ to determine the fraction of cargo-targeted spines and the frame of spine entry. The mean percentage of spines targeted with peroxisomes was compared with a Mann–Whitney test (Fig. 2i), and the mean percentages of peroxisome and endosome entries were subjected to a Kruskal–Wallis ANOVA with Dunn's post-hoc test (Fig. 2j) or a one-way ANOVA with Bonferroni's post-hoc test (Extended Data Fig. 4p), respectively. The half-time of peroxisome entries into dendritic spines after illumination was estimated by fitting a single exponential function ($R^2 = 0.9942$) through the inverted cumulative histogram of the observed entry events after 491 nm illumination was stopped.

Analysis of axon growth and growth cone area. Axonal growth was manually tracked using the MTrackJ plugin in ImageJ[37]. The percentages of growth cones exhibiting light-induced reduction in dynamics or growth enhancements were compared using unpaired two-tailed t-tests (Fig. 3g, h). In all our experiments, only the RFP channel was available for imaging without triggering photo-heterodimerization before, during and after exposure to blue light. We used this channel to image FKBP-tagRFPt-RAB11 to verify that light-controlled recruitment of BICDN induced the removal of RAB11 endosomes (Extended Data Fig. 8a, b). FKBP-RAB11 was enriched at vesicles-like structures, whose dynamics altered upon light-dependent recruitment of BICDN to recruit dynein. In addition, FKBP-RAB11 diffusely filled the axon, which could be used to determine axon morphology and size with precision comparable to a cytoplasmic GFP fill (Extended Data Fig. 8c). We counted the positive pixels in a binarized image obtained by thresholding the median-filtered tagRFPt image, followed by two erosions and closure[38]. Because tag-RFPt fluorescence of this construct showed a threefold increase in intensity upon 491 nm excitation (see Extended Data Fig. 8a), we established a dynamic threshold T that scaled with the maximum intensity of the object, that is, $T = I_{bg} + \sigma_{bg} + 0.02(I_{max})$, in which I_{bg} and σ_{bg} are the average and standard deviation of the intensity in an area outside the axon, respectively, and I_{max} is the average of the top 2% intensity values above $I_{bg} + \sigma_{bg}$. Using these parameters, changes in area are independent of the changes in intensity upon exposure to blue light or due to dynein-mediated removal of RAB11 vesicles (see Extended Data Fig. 8d, e).

Relative decreases in growth cone RAB11-FKBP signal were calculated by rescaling all intensity values normalized initially to $t_{-2:30\,min}$ to the average intensity value of –FRB–LOV control growth cones at $t_{8\,min}$ (see Extended Data Fig. 8a). To calculate changes in growth cone area before blue-light illumination, we compared single growth cone area values averaged over three frames at $t_{0\,min}$ and $t_{-4\,min}$ relative to illumination onset (See Extended Data Fig. 8g, h). Analogously, comparing values at $t_{8\,min}$ and $t_{0\,min}$ shows net growth during blue-light illumination (see Extended Data Fig. 8i, j). All of these results were compared using Mann–Whitney tests (Extended Data Fig. 8b, g, i). All statistical testing was performed in GraphPad Prism 5 software. No statistical method was used to predetermine sample size.

**Immunofluorescence cell staining, imaging and antibodies.** COS-7 cells (1 day after transfection) or primary hippocampal neurons (2 days after transfection) were either kept in the dark or illuminated for 10 min using a blue light-emitting diode mounted in the incubator. Afterwards, cells were fixed at room temperature for 10 min with 4% paraformaldehyde (PFA), 4% sucrose. For detection of EB1, cells were fixed for 5 min at $-20\,°C$ in 100% ice-cold methanol supplemented with 1 mM EGTA, followed by 5 min post-fixation at room temperature in 4% PFA, 4% sucrose. After fixation, cells were washed three times in PBS and incubated overnight at $4\,°C$ in GDB buffer (0.1% BSA, 450 mM NaCl, 0.3% Triton X-100 and 16.7 mM phosphate buffer, pH 7.4) containing the primary antibody. The next day, cells were washed three times for 10 min with PBS, followed by a 1-h incubation at room temperature with the secondary antibody in GDB buffer. After washing cells three times for 10 min in PBS, slides were mounted in Vectashield mounting medium (Vector Laboratories). Images were taken with a Nikon eclipse 80i upright fluorescence microscope and a Coolsnap HQ2 CCD camera (Photometrics), using a $40\times$ oil objective (Plan Fluor, NA 1.3), $60\times$ oil objective (Plan Apo VC, NA 1.4) or $100\times$ oil objective (Plan Apo VC, NA 1.4).

Antibodies and reagents used: mouse anti-Cytochrome c (6H2.B4, 556432, BD Biosciences), mouse anti-PDI (RL90, MA3-019, Affinity BioReagents), phalloidin-Alexa647 (A22287, Invitrogen), mouse anti-alpha tubulin (B-5-1-2, T-5168, Sigma), mouse anti-EB1 (610535, BD Transduction), mouse anti-Lamp1 (this antibody developed by J. T. August and J .E. K. Hildreth, was obtained from the hybridoma bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology), mouse anti-EEA1 (BD Transduction), rabbit anti-RAB11 (71-5300, Invitrogen), rabbit anti-Homer-1 (160-002, SySy), Alexa 488-, Alexa 568-, Alexa 647-conjugated secondary antibodies (Invitrogen).

GFP pull-down and western blotting. HEK cells were cultured in DMEM/Ham's F10 (1:1) medium containing 10% FCS and penicillin/streptomycin. Then 1 day after plating, HEK293T cells were transfected using polyethylenimine (PEI; Polysciences). After 24 h, GFP beads (GFP-Trap_M, Chromotek) were washed in washing buffer (TBS, 0.5% Triton X-100 and protease inhibitor) and incubated for 1 h in blocking buffer (TBS, 0.5% Triton X-100, 2% glycerol, 2% chicken egg white). Meanwhile, cells were collected in ice-cold TBS, pelleted and lysed in extraction buffer (TBS 0.5% Triton X-100, protease inhibitor, phosphatase inhibitor (Roche), 100 μM GTPγS, 5 mM MgCl$_2$, pH 8.0). Cell lysates were centrifuged for 15 min at $4\,°C$ at 12,000g, followed by a 1.5-h incubation of the supernatants with the washed GFP beads, while rotating at $4\,°C$. Beads were collected using a magnetic separator and washed four times. Samples were eluted in SDS sample buffer, boiled for 5 min and loaded onto SDS–PAGE gels and blotted on PVDF membranes (Millipore). Blots were blocked in 5% milk in PBST (0.1% Tween in PBS) and incubated overnight at $4\,°C$ (primary antibody) or for 1 h at room temperature (secondary antibody conjugated to horseradish peroxidase) in PBST. Finally, blots were developed using enhanced chemiluminescent western blotting substrate (Pierce).

Antibodies used: rabbit anti-TagRFPt (ab234, Evrogen), rabbit anti-GFP (ab290, abcam) and anti-rabbit IgG antibody conjugated to horseradish peroxidase (DAKO).

29. Shaner, N. C. et al. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nature Methods* **5,** 545–551 (2008).
30. Miyamoto, T. et al. Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nature Chem. Biol.* **8,** 465–470 (2012).
31. Hoogenraad, C. C. et al. Neuron specific Rab4 effector GRASP-1 coordinates membrane specialization and maturation of recycling endosomes. *PLoS Biol.* **8,** e1000283 (2010).
32. Peden, A. A. et al. The RCP-Rab11 complex regulates endocytic protein sorting. *Mol. Biol. Cell* **15,** 3530–3541 (2004).
33. Schlager, M. A. et al. Pericentrosomal targeting of Rab6 secretory vesicles by Bicaudal-D-related protein 1 (BICDR-1) regulates neuritogenesis. *EMBO J.* **29,** 1637–1651 (2010).
34. Honnappa, S. et al. An EB1-binding motif acts as a microtubule tip localization signal. *Cell* **138,** 366–376 (2009).
35. De Paola, V., Arber, S. & Caroni, P. AMPA receptors regulate dynamic equilibrium of presynaptic terminals in mature hippocampal networks. *Nature Neurosci.* **6,** 491–500 (2003).
36. Kapitein, L. C., Yau, K. W. & Hoogenraad, C. C. Microtubule dynamics in dendritic spines. *Methods Cell Biol.* **97,** 111–132 (2010).
37. Meijering, E., Dzyubachyk, O. & Smal, I. Methods for cell and particle tracking. *Methods Enzymol.* **504,** 183–200 (2012).
38. Russ, J. C. *The Image Processing Handbook* 5th edn (CRC/Taylor and Francis, 2007).
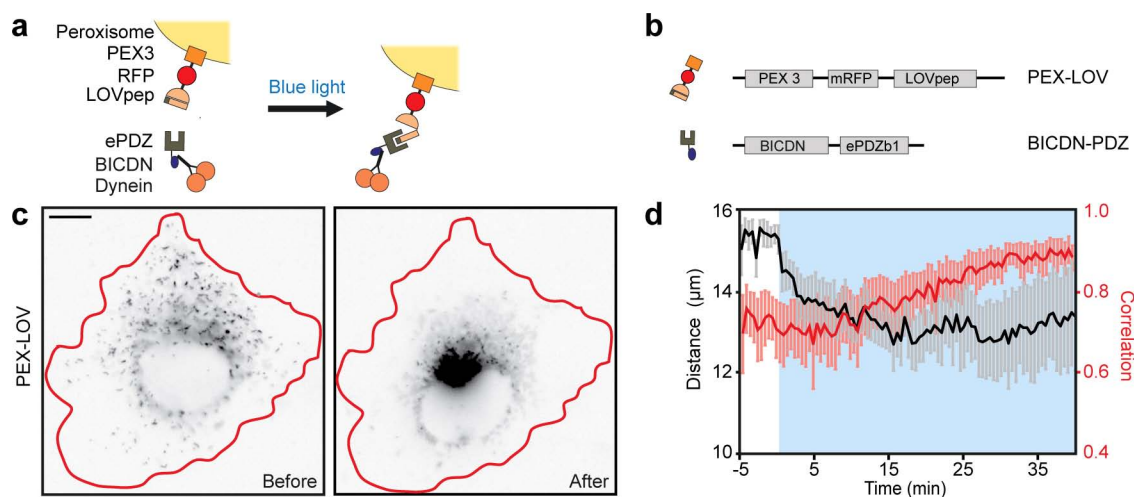
Extended Data Figure 1 | Optical control of dynein-based cargo motility. **a**, **b**, Assay and constructs. A fusion construct of PEX3, monomeric red fluorescent protein (mRFP) and LOVpep (PEX–LOV) targets peroxisomes. After blue-light illumination, a fusion of the N terminus of the dynein adaptor BICD2 and ePDZb1 (BICDN–PDZ) is recruited to peroxisomes. **c**, Peroxisome distribution in a COS-7 cell expressing PEX–LOV and BICDN–PDZ before and during light-induced recruitment of dynein (inverted contrast). Red lines indicate cell outline. Scale bar, 10 μm. See Supplementary Video 1. **d**, Black: time trace of $R_{90\%}$ (radius of circle enclosing 90% of cellular fluorescence; see Methods) in cells expressing PEX–LOV and BICDN–PDZ ($n = 5$ cells). Red: correlation index (frame-to-frame differences in the peroxisome recordings; see Methods) of the same cells. Blue-light illumination is indicated in blue; mean ± s.e.m.
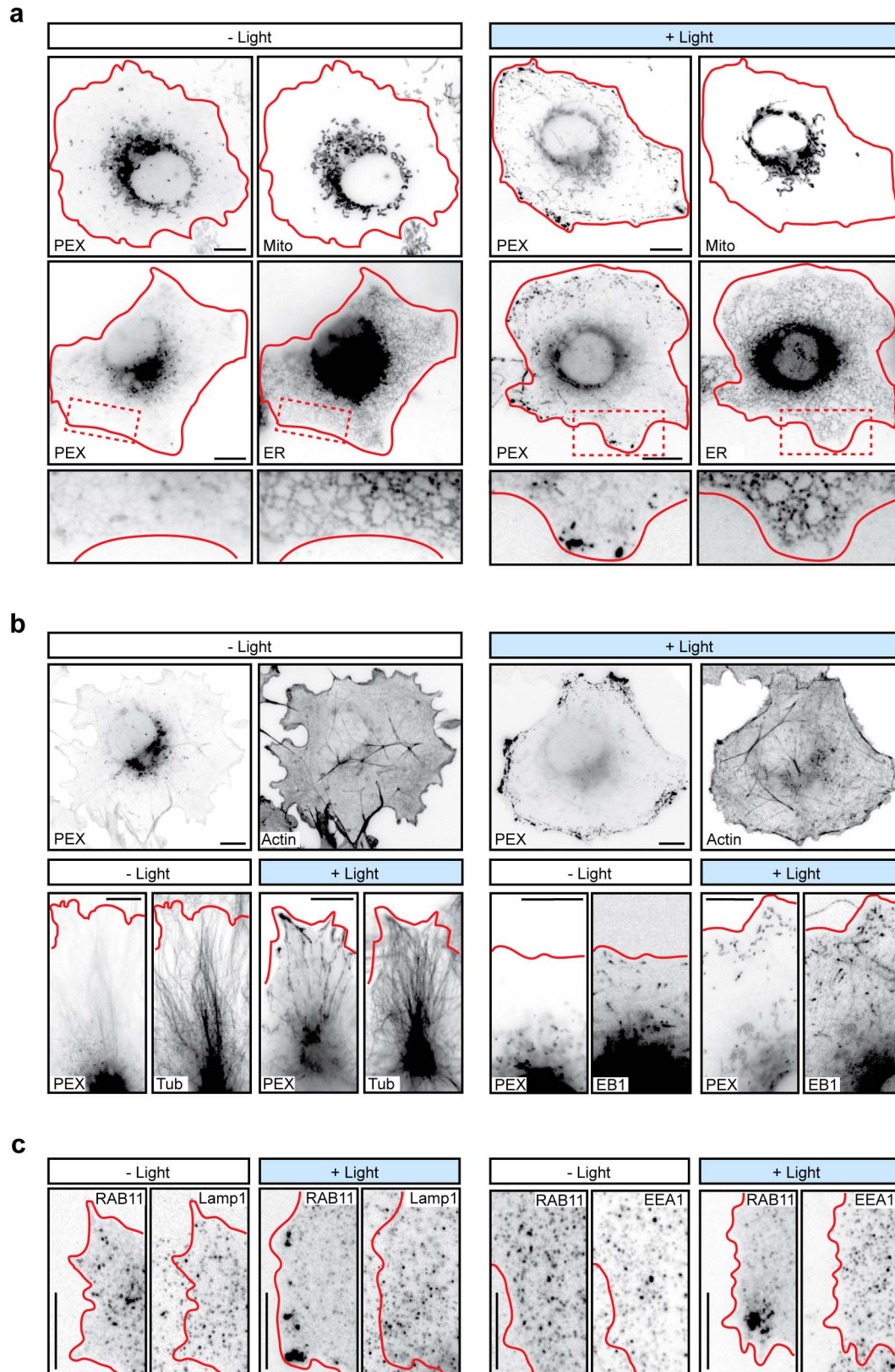
Extended Data Figure 2 | Light-induced organelle redistribution is organelle-specific and does not affect the cytoskeleton. a, Images of fixed cells expressing PEX–LOV and KIF–PDZ, showing the distribution of peroxisomes and mitochondria (anti-cytochrome-c), or peroxisomes and the endoplasmic reticulum (anti-protein disulfide isomerase (PDI)) in the absence (left) or presence (right) of blue light. **b**, Images of fixed cells expressing PEX–LOV and KIF–PDZ, showing the distribution of peroxisomes and phalloidin, α-tubulin or EB1 staining in the absence or presence of blue light. **c**, Images of fixed cells expressing FKBP–RAB11, FRB–LOV and KIF–PDZ, showing the distribution of RAB11 recycling endosomes together with lysosomes (anti-Lamp1) or early endosomes (anti-EEA1) in the absence or presence of blue light. Red lines indicate cell outline. Scale bars, 10 μm.
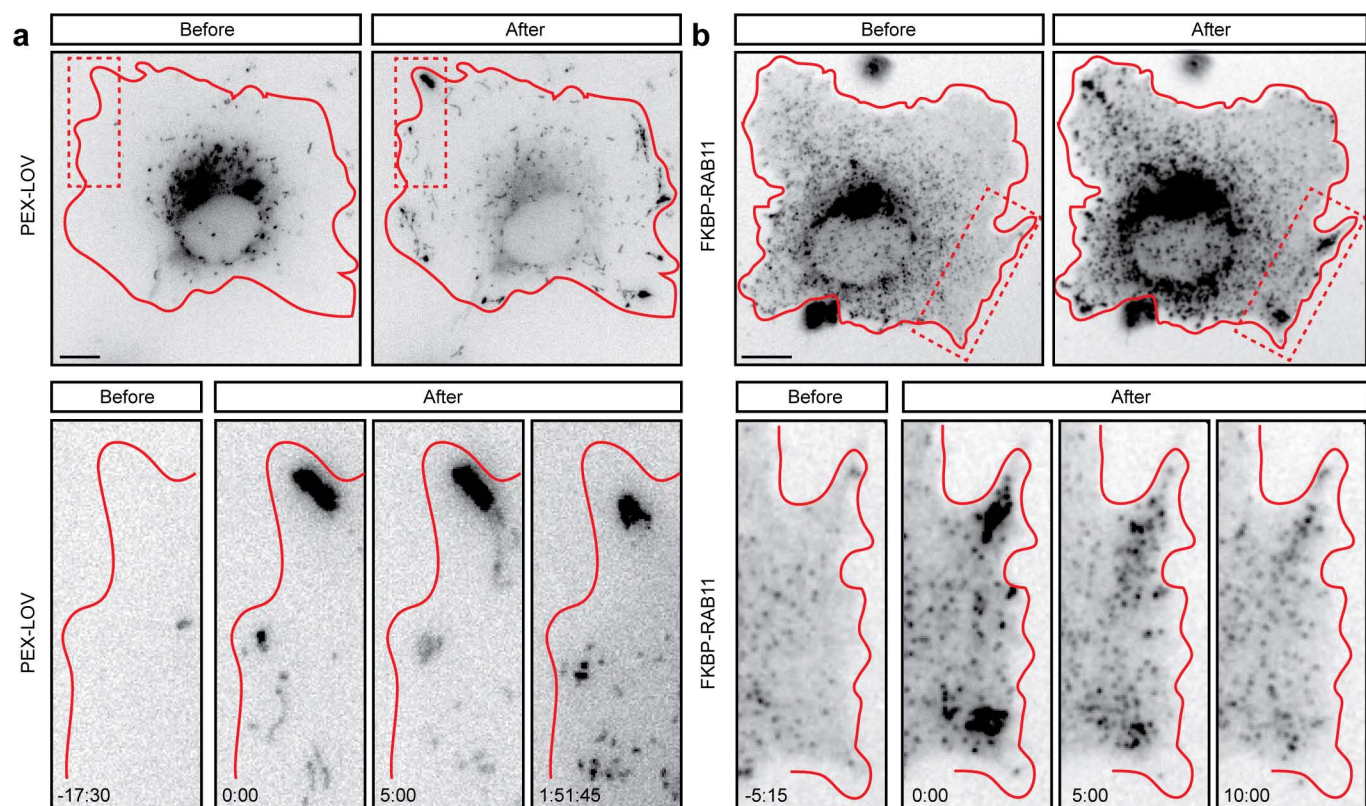
Extended Data Figure 3 | After light induced organelle displacement, peroxisomes remain at their newly obtained position whereas the distribution of recycling endosomes quickly reverses back to normal. **a**, Peroxisome distribution before and after exposure to blue light in cells expressing PEX–LOV and KIF–PDZ. Blue-light illumination was terminated at $t_{0:00}$. **b**, Distribution of RAB11 recycling endosomes before and after exposure to blue light in cells expressing FKBP–RAB11, FRB–LOV and KIF–PDZ. Blue light was turned off at $t_{0:00}$. Red lines indicate cell outline. Scale bars, 10 μm. See Supplementary Video 4.
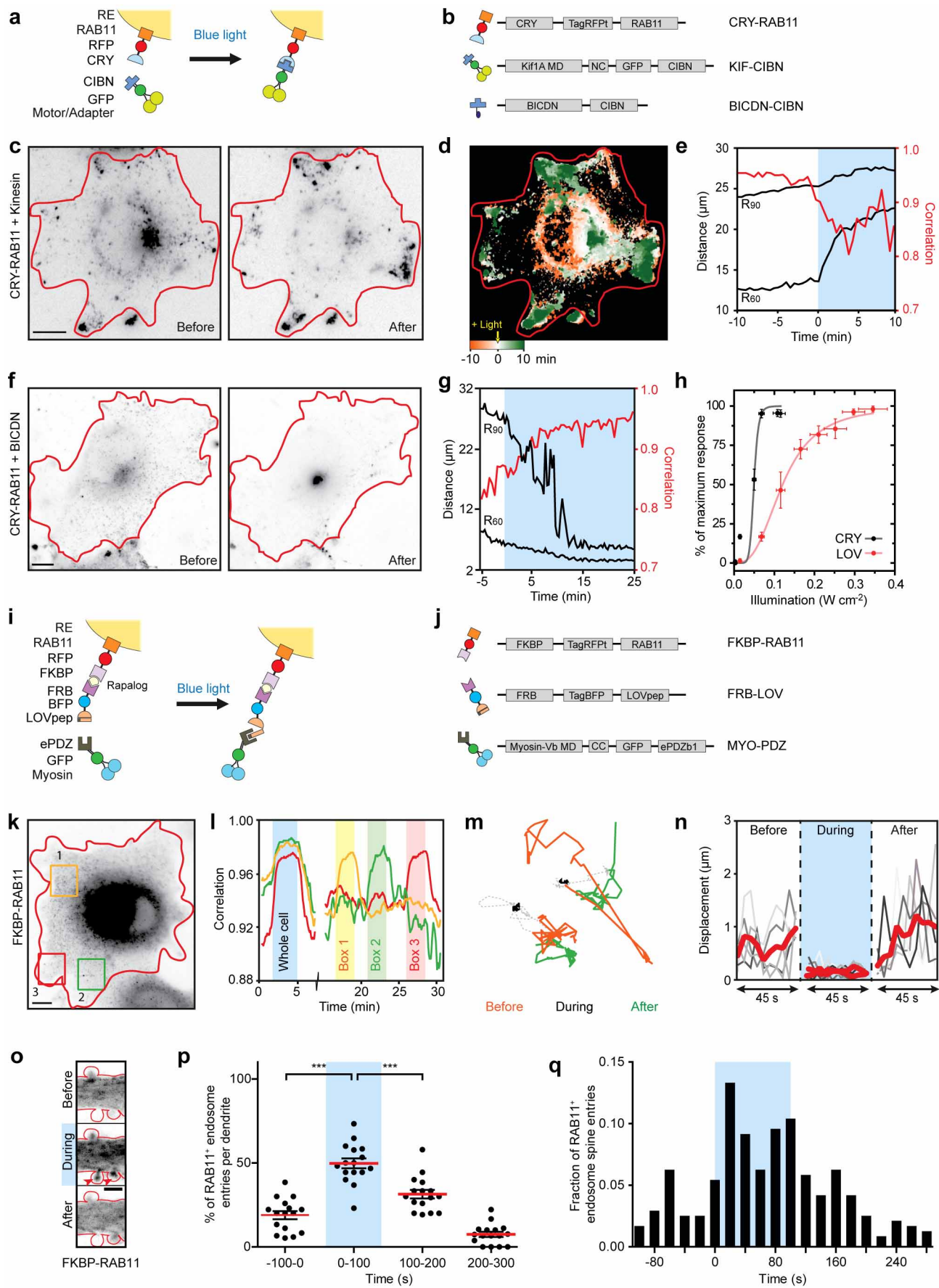
Extended Data Figure 4 | Spatiotemporal control of recycling endosome distribution and dynamics. a, b, Assay and constructs. A fusion construct of Cry2PHR, tagRFPt and RAB11, (CRY–RAB11) targets RAB11 recycling endosomes. After blue-light illumination, a fusion of truncated KIF1A, GFP and CIBN (KIF–CIBN) or a fusion of truncated BICDN, GFP and CIBN (BICDN–CIBN) is recruited to RAB11 recycling endosomes. **c**, RAB11 vesicle distribution before and after light-induced recruitment of KIF1A (inverted contrast). Red lines indicate cell outline. Scale bar, 10 μm. **d**, Overlay of sequential binarized images from the recording in **c**, colour-coded by time as indicated. Orange marks the initial distribution of RAB11 vesicles, whereas green marks regions targeted after exposure to blue light. **e**, Time trace of the $R_{60\%}$ and $R_{90\%}$ (black) and the correlation index (red) of the cell shown in **c** and **d**. Blue box marks blue-light illumination. **f**, RAB11 distribution in a cell expressing CRY–RAB11 and BICDN–CIBN before and after blue-light illumination (inverted contrast). Red lines indicate cell outline. Scale bar, 10 μm. **g**, Time trace of the $R_{60\%}$ and $R_{90\%}$ (black) and correlation index for the cell shown in **f**. **h**, Irradiance response curve for cells transfected with CRY–PEX and KIF–CIBN (red), or PEX–LOV plus KIF–PDZ (black). To exclude activation failure due to poorly expressed motors, the number of cells reacting at each concentration was divided by the number of cells responding to subsequent high irradiance ($\sim$1.3 W cm^{-2}). Three biological replicates. Cells per intensity (for increasing intensities): 28, 21, 22, 20, 24, 22 and 20 for CRY, 30, 28, 33, 31, 28, 33, 33, 32 and 26 for LOV. Error bars depict s.e.m.; three biological replicates. Solid line shows fit to $R = 100 \cdot I^n / \left(I_0^n + I^n\right)$, with R the response, I the illumination intensity, I_0 the intensity at which the response is 50%, and n the Hill coefficient. For CRY–PEX and PEX–LOV, I_0 is 0.05 and 0.12 W cm^{-2}, respectively. **i, j**, Assay and constructs. A fusion construct of FKBP, tagRFPt and RAB11 (FKBP–RAB11) targets RAB11 recycling endosomes. Rapalog addition couples FKBP to FRB, leading to recruitment of the FRB, tagBFP and LOVpep fusion protein (FRB–LOV). After blue-light illumination a fusion of truncated myosin-Vb, GFP and ePDZb1 (MYO–PDZ) is recruited to RAB11 vesicles. **k**, RAB11 distribution in a cell expressing FKBP–RAB11, FRB–LOV and MYO–PDZ before sequential blue-light illumination of the regions marked with numbered boxes (inverted contrast). Scale bar, 10 μm. See Supplementary Video 5. **l**, Time traces of the correlation index in the areas shown in **k**. Blue box marks whole-cell exposure to blue light, whereas colored boxes indicate local illumination. **m**, Example trajectories of two RAB11 recycling endosomes before, during and after recruitment of myosin-Vb, as indicated. Data was acquired with 1-s intervals. For each period 40 s are shown. **n**, Frame-to-frame displacements of RAB11 recycling endosomes before, during and after light-induced recruitment of myosin-Vb (5 s interval). Thick lines show the average of five tracks in shades of grey. **o**, FKBP–RAB11 distribution (inverted contrast) in a dendrite and dendritic spines before, during and after blue-light illumination. Images are maximum projections spanning 60 s. Red lines indicate cell outline, arrowheads mark spines targeted with recycling endosomes during blue-light illumination. Scale bar, 2 μm. **p**, Percentage of recycling endosome spine entry events per dendrite before, during and after illumination in bins of 100 s. Blue box indicates blue-light illuminated interval, $n = 16$ dendrites in three independent experiments. Red bar denotes mean $\pm$ s.e.m., ***$P < 0.0001$, one-way ANOVA, Bonferroni's post-hoc test. **q**, Histogram of fraction of all ($n = 237$) recycling endosome spine entries in bins of 20 s. Blue box indicates blue-light-illuminated interval.

a

b

c

d

e
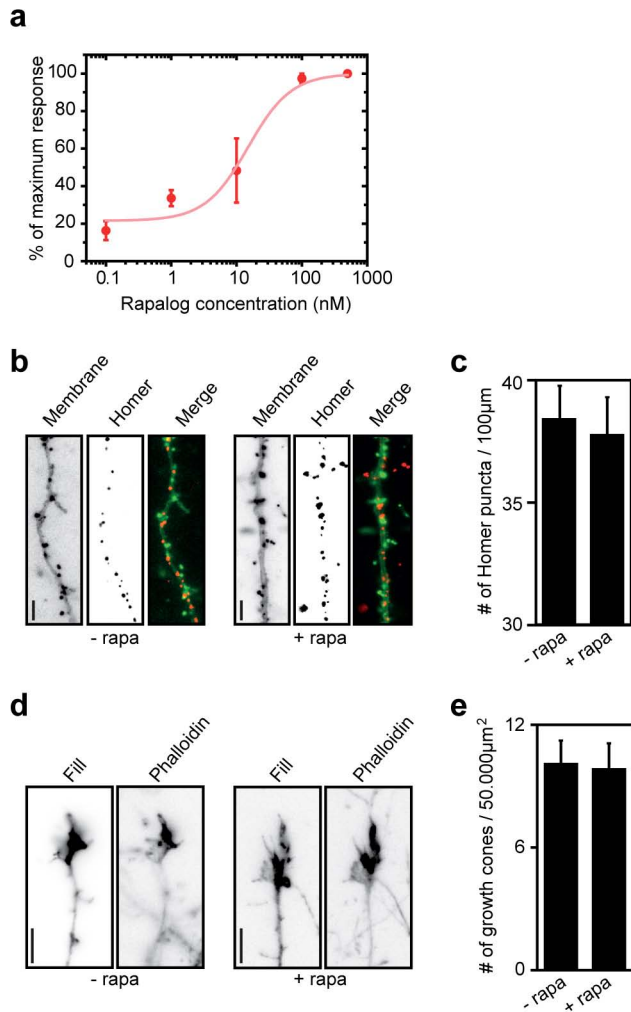
Extended Data Figure 5 | Rapalog in the nanomolar range is sufficient to recruit FRB–LOV to FKBP–RAB11 and does not affect the number of spines or growth cones in hippocampal neurons. a, Response curve of RAB11 recycling endosome relocalization in cells expressing FKBP–RAB11, FRB–LOV and KIF–PDZ exposed to blue light in relation to rapalog concentration. To exclude activation failure due to poorly expressed motors, the number of cells reacting at each concentration was divided by the number of cells responding to subsequent high rapalog concentration ($1 \mu M$). Solid line shows fit to $R = \left(R_{min}I_0^n + 100 \cdot I^n\right) / \left(I_0^n + I^n\right)$, with R the response, c the rapalog concentration, c_0 the concentration at which the response is 50%, n the Hill coefficient, and R_{min} the response at 0 mM rapalog. R_{min} is 22% and c_0 is 15 nM. $n = 30$ (0.1 nM), 37 (1 nM), 30 (10 nM), 28 (100 nM) and 28 (500 nM) responsive cells from three independent experiments. Error bars depict s.e.m. **b,** Hippocampal neurons transfected with membrane–GFP incubated for 2.5 h in the presence or absence of 100 nM rapalog, co-stained with the post-synaptic marker Homer. **c,** Quantification of the number of Homer puncta per $100 \mu m$ dendrite length in the presence or absence of 100 nM rapalog ($n = 13$ neurons per condition). Error bars depict s.e.m. **d,** Hippocampal neurons transfected with GFP incubated for 2.5 h in the presence or absence of 100 nM rapalog, co-stained with phalloidin. **e,** Quantification of the number of growth cones per $50{,}000 \mu m^2$ in the presence or absence of 100 nM rapalog, co-stained with phalloidin. $n = 19$. Scale bars, $5 \mu m$. Error bars depict s.e.m.
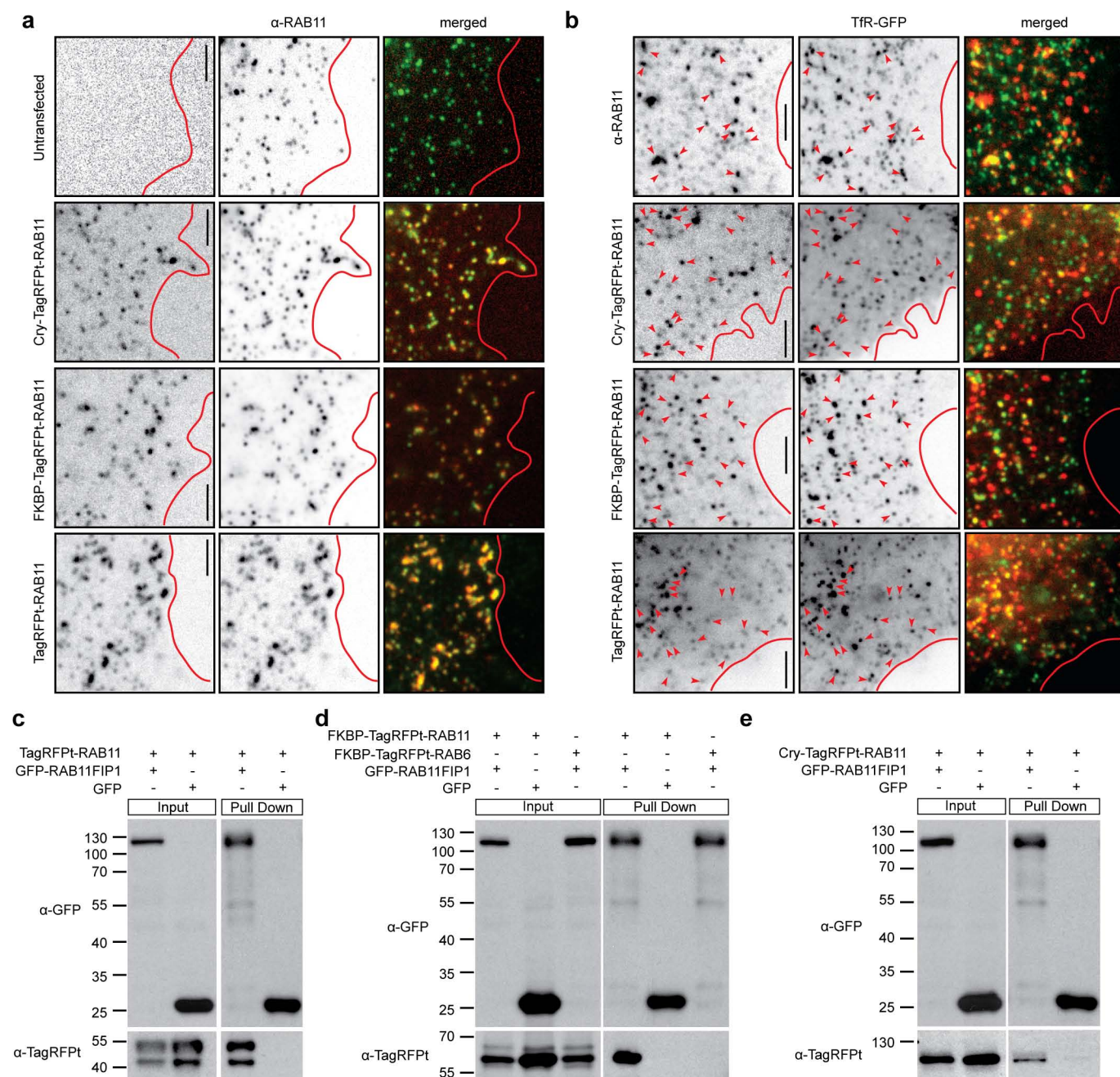
Extended Data Figure 6 | RAB11 fusion constructs are recognized by the RAB11 antibody, partially co-localize with transferrin receptors and interact with RAB11FIP1. **a**, Images of untransfected cells or cells transfected with CRY–RAB11, FKBP–RAB11 or tagRFPt–RAB11, co-stained with anti-RAB11 antibody (inverted contrast). Red lines indicate cell outline. Scale bar, 2.5 μm. **b**, Images of cells transfected with TfR-GFP only, or co-transfected with CRY–RAB11, FKBP–RAB11 or tagRFPt–RAB11 (inverted contrast). Red lines indicate cell outline. Scale bar, 2.5 μm. **c–e**, GFP pull-down assays with lysates of HEK cells expressing GFP or GFP–RAB11FIP1 together with tagRFPt–RAB11 (**c**), FKBP–RAB11 or FKBP–tagRFPt–RAB6 (**d**) or CRY–RAB11 (**e**) were analysed by western blotting using antibodies against tagRFPt and GFP.
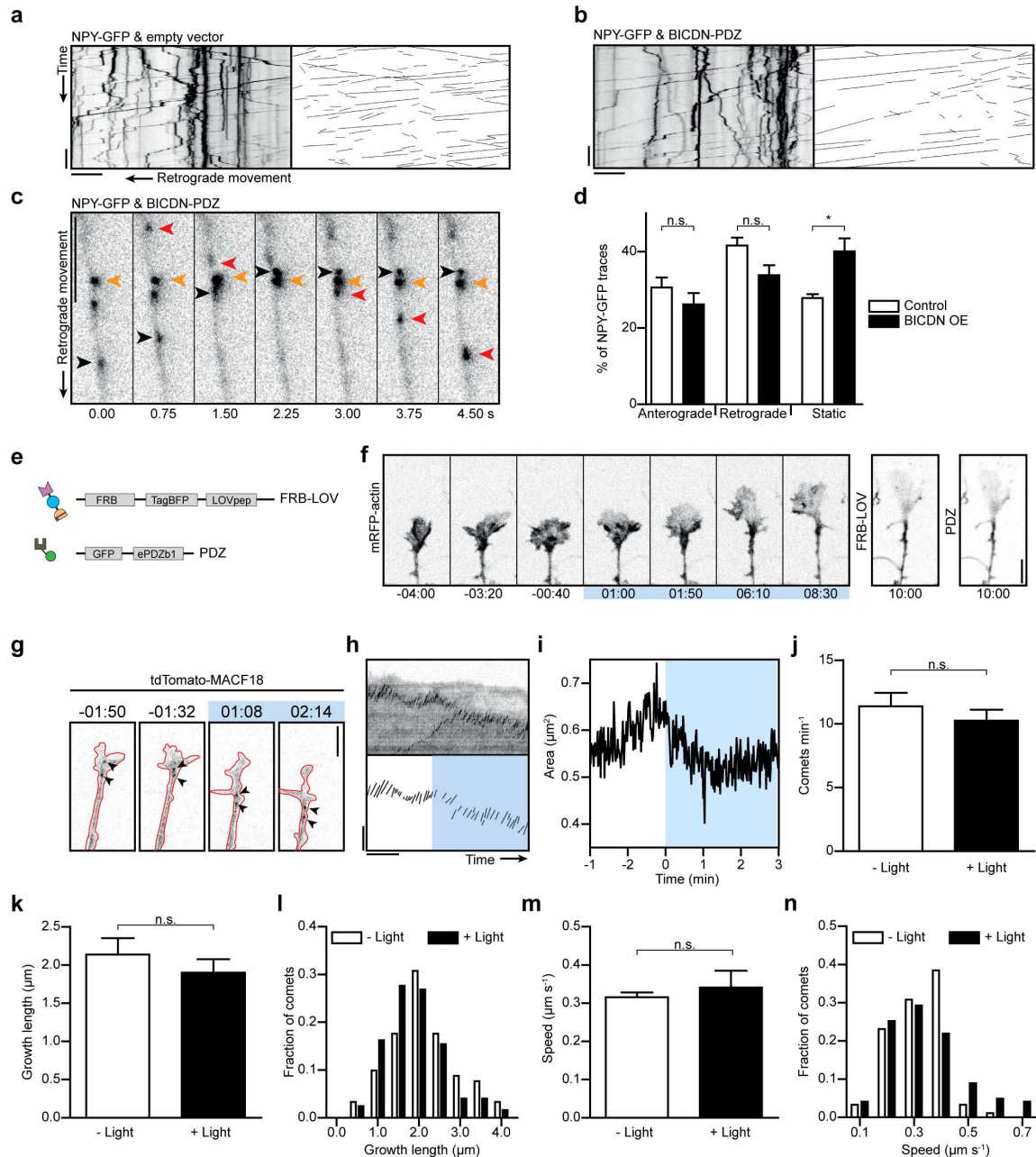
Extended Data Figure 7 | BICDN overexpression does not significantly inhibit dynein-based transport and the growth cone cytoskeleton is not affected by light-induced recruitment of BICDN to recycling endosomes. a, b, Left: kymograph of dense-core vesicles motility in an axon expressing neuropeptide Y (NPY) fused to GFP (NPY–GFP) and empty vector (a) or BICDN–PDZ (b) (inverted contrast), representative of $n = 5$ and $n = 10$ axons, respectively. Right: corresponding binary image of traces used for further analysis of anterograde and retrograde movements. Scale bars, 5 µm and 10 s. c, Position of dense-core vesicles along an axon expressing NPY–GFP and BICDN–PDZ. Single coloured arrowheads point to the same vesicle, highlighting retrograde (red), anterograde (black) and non-moving (orange) vesicles. Scale bar, 5 µm. d, Quantification of the percentage of static, anterograde and retrograde moving vesicles from kymographs shown in a and b in axons with ($n = 10$) or without ($n = 5$) BICDN–PDZ overexpression. Graph shows mean ± s.e.m., *$P > 0.05$, one-way ANOVA and Bonferroni's multiple comparison test. e, A fusion of FRB, tagBFP and LOVpep (FRB–LOV) and a fusion of GFP and ePDZb1 (PDZ) were expressed in neurons. After blue-light illumination, LOVpep undergoes a conformational change, allowing binding of PDZ to FRB–LOV. f, Actin dynamics in growth cones coexpressing mRFP–actin along with the constructs shown in e, in response to light-induced heterodimerization of FRB–LOV and PDZ, representative of $n = 5$ growth

cones. The blue box indicates the interval of blue-light illumination. Scale bar, 5 µm. g, Imaging of growing microtubule (MT) plus ends using tdTomato-MACF18 shows the dynamics of microtubule plus-ends in growth cones before and during blue-light illumination in neurons co-expressing FKBP–RAB11, FRB–LOV and BICDN–PDZ. Red line indicates cell outline, arrowheads point at plus-ends. Scale bar, 5 µm. h, Kymograph of MACF18 comets of the growth cone shown in g and binarized traces used for analysis, representative of $n = 4$ growth cones. Blue box indicates blue-light illumination interval. Scale bars, 5 µm and 1 min. i, Area measurement of growth cone shown in g before and during blue-light illumination. j, Quantification of the number of MACF18 comets per minute in growth cones before and during blue-light illumination ($n = 4$ neurons). Graph shows mean ± s.e.m. Paired two-tailed t-test, $n = 4$ cells. k, Quantification of the growth length of MACF18 comets in growth cones before and during blue-light illumination ($n = 4$ neurons). Graph shows mean ± s.e.m. Paired two-tailed t-test, $n = 4$ cells. l, Distribution of fraction of MACF18 comets per growth length in bins of 0.5 µm ($n = 214$ traces). m, Quantification of the growth speed of MACF18 comets in growth cones before and during blue-light illumination ($n = 4$ neurons). Graph shows mean ± s.e.m. Paired two-tailed t-test, $n = 4$ cells. n, Distribution of fraction of MACF18 comets per growth speed in bins of 0.1 µm s^{-1} ($n = 214$ traces).
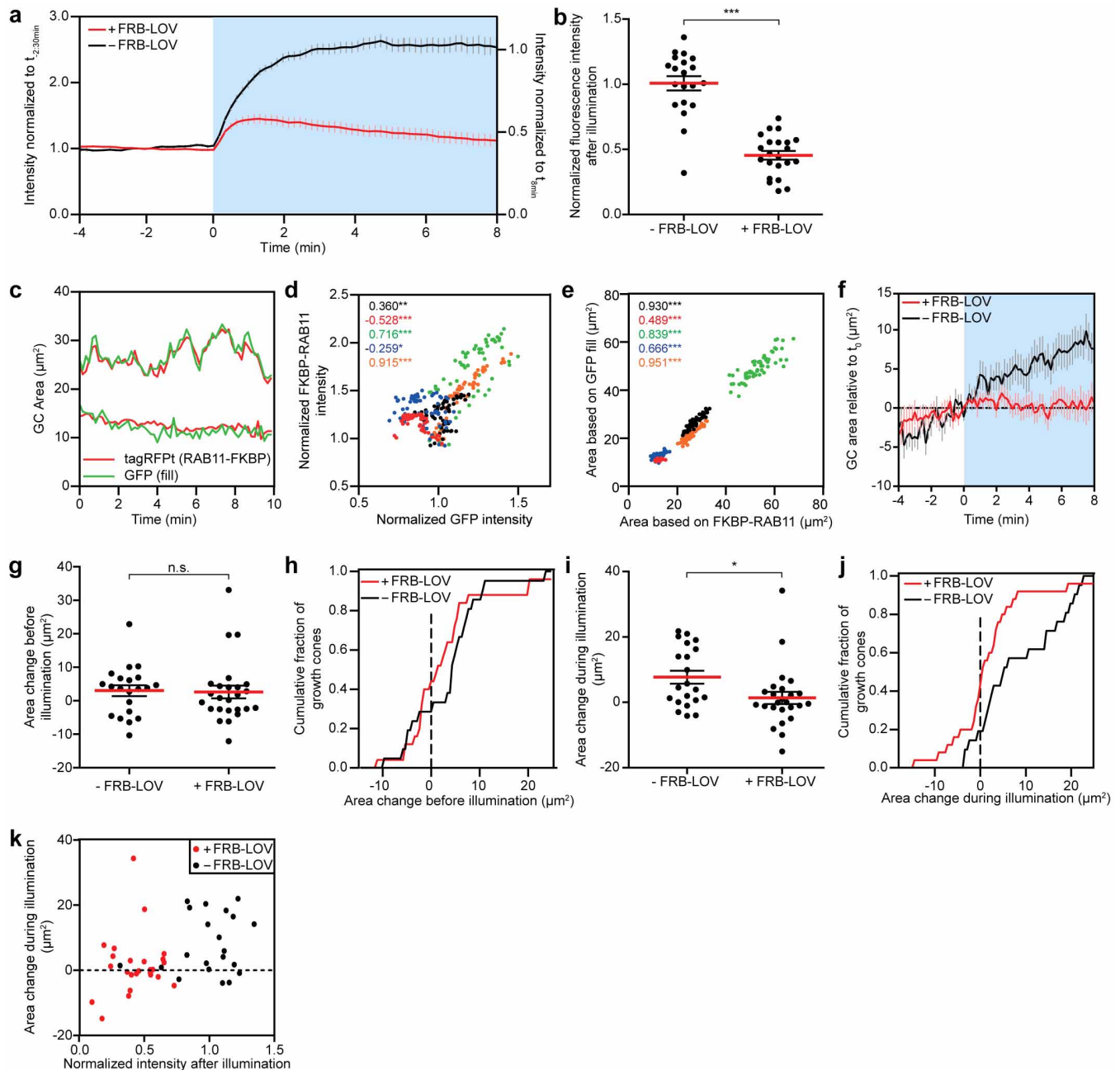
Extended Data Figure 8 | Intensity rescaling and accurate growth cone area measurements based on RAB11 fluorescence. a, Mean intensity of growth cone FKBP–RAB11 fluorescence from neurons expressing BICDN–PDZ in the absence (black, $n = 21$) or presence of FRB–LOV (red, $n = 25$) normalized to the intensity before ($t_{-2:30\ min}$) (left axis) and rescaled relatively to the intensity of –LOV growth cones at $t_{8\ min}$ (right axis). Blue box indicates blue-light illuminated interval. Graph shows mean ± s.e.m. **b**, Quantification of FKBP–RAB11 fluorescence intensity in **a** after 8 min of blue-light illumination, normalized to the average fluorescence at $t_{8\ min}$ in control neurons. Graph shows mean ± s.e.m., ***$P < 0.0001$, Mann–Whitney test. **c**, Area measurements of two representative growth cones from neurons expressing FKBP–RAB11, FRB–LOV, BICDN–PDZ and soluble GFP over time. Representative of five growth cones (shown in **d** and **e**). **d**, Normalized tagRFPt-RAB11 intensity of five growth cones as in **c** plotted against their normalized GFP intensity. Intensity values are averaged over the first five frames per growth cone. Pearson correlation coefficient (r) for each growth cone is indicated in top left corner. Same colour indicates measurements of the same growth cone. **e**, FKBP–RAB11-based area measurements plotted against GFP-based area measurements of the same growth cones as in **d**. Pearson correlation coefficient (r) for each growth cone is indicated in top left corner. Same colour indicates measurements of the same growth cone. **f**, Traces of

growth cone area measurements based on FKBP–RAB11 signal in the absence ($n = 25$, red trace) and presence of FRB–LOV ($n = 21$, black trace) in growth cones before and during blue-light illumination (see Methods). Graph shows mean ± s.e.m. Blue box indicates blue light-exposed interval. **g**, Quantification of the area increase in the absence and presence of FRB–LOV in growth cones during blue-light illumination (−4 to 0 min). Values per growth cone are averaged over three frames. Graph shows mean ± s.e.m. $P = 0.4145$ (n.s., not significant), Mann–Whitney test. **h**, Cumulative histogram showing the fraction of growth cones with area shrinkage or growth (left or right of dashed line, respectively) before blue-light illumination (−4 to 0 min). Values per growth cone are averaged over three frames. **i**, Quantification of the area change of −FRB–LOV and +FRB–LOV growth cones during blue-light illumination (0 to 8 min). Values per growth cone are averaged over three frames. Graph shows mean ± s.e.m., *$P = 0.0206$, Mann–Whitney test. **j**, Cumulative histogram showing the fraction of growth cones with area shrinkage or growth (left or right of dashed line, respectively) during blue-light illumination (0–8 min). Values per growth cone are averaged over three frames. **k**, Scatter plot showing net growth during blue-light illumination and normalized fluorescence intensity after blue-light illumination per +FRB–LOV (red) or −FRB–LOV (black) growth cone.

LETTER

Metabolic coupling of two small–molecule thiols programs the biosynthesis of lincomycin A

Qunfei Zhao[1]*, Min Wang[1]*, Dongxiao Xu[1], Qinglin Zhang[2] & Wen Liu[1,2]

Low-molecular-mass thiols in organisms are well known for their redox-relevant role in protection against various endogenous and exogenous stresses[1–3]. In eukaryotes and Gram-negative bacteria, the primary thiol is glutathione (GSH), a cysteinyl-containing tripeptide. In contrast, mycothiol (MSH), a cysteinyl pseudo-disaccharide, is dominant in Gram-positive actinobacteria, including antibiotic-producing actinomycetes and pathogenic mycobacteria. MSH is equivalent to GSH, either as a cofactor or as a substrate, in numerous biochemical processes[4], most of which have not been characterized, largely due to the dearth of information concerning MSH-dependent proteins. Actinomycetes are able to produce another thiol, ergothioneine (EGT), a histidine betaine derivative that is widely assimilated by plants and animals for variable physiological activities[5]. The involvement of EGT in enzymatic reactions, however, lacks any precedent. Here we report that the unprecedented coupling of two bacterial thiols, MSH and EGT, has a constructive role in the biosynthesis of lincomycin A, a sulfur-containing lincosamide (C8 sugar) antibiotic that has been widely used for half a century to treat Gram-positive bacterial infections[6–9]. EGT acts as a carrier to template the molecular assembly, and MSH is the sulfur donor for lincomycin maturation after thiol exchange. These thiols function through two unusual S-glycosylations that program lincosamide transfer, activation and modification, providing the first paradigm for EGT-associated biochemical processes and for the poorly understood MSH-dependent biotransformations, a newly described model that is potentially common in the incorporation of sulfur, an element essential for life and ubiquitous in living systems.

Mycothiol (MSH, Fig. 1a) mediated detoxification typically relies on the conjugation of MSH to an electrophilic toxin. An amidase, Mca, then hydrolyses the resulting MSH S-conjugate to produce a pseudo-disaccharide unit, 1-O-glucosamine-D-myo-inositol (GlcN-Ins), and a mercapturic acid derivative, which is an excretive N-acetyl-cysteinyl product common in GSH-mediated metabolism[10–12] (Fig. 1b). In actinomycetes, mca orthologues have been found in several biosynthetic gene clusters of antibiotics[4,11], including that of lincomycin A, suggesting that MSH S-conjugation is associated with the production of these antibiotics. Lincomycin A consists of an N-methylated 4-propyl-L-proline (PPL) moiety and lincosamide, an unusual eight-carbon sugar decorated with a methylmercapto group at C-1 (Fig. 1d). Cell protection against the activity of lincomycin A depends largely on methylation of the bacterial ribosome, whereby the molecule mimics the 3′ end of (de)acetyl-tRNA and blocks protein synthesis at the initial stage of the elongation cycle[13,14]. This fact, in combination with the methylmercapto group found in the structure, leads to the proposal that MSH S-conjugation has a constructive role in lincomycin biosynthesis by supplying sulfur rather than a protective role in antibiotic detoxification. To validate this hypothesis, we inactivated lmbE, a mca orthologue that is located in the lincomycin biosynthetic gene cluster (lmb) in Streptomyces lincolnensis[15,16] (Extended Data Fig. 2a).

As anticipated, the ΔlmbE mutant strain accumulated a MSH-associated lincomycin analogue, 1 (Fig. 2). In this molecule, the C8-sugar unit is appended to MSH via an α-S-linkage (Supplementary Text), identical to that of lincomycin A in configuration. The ΔlmbE mutant strain still produced lincomycin A, albeit in a lower yield, indicating that an additional mca orthologue is present outside the lmb cluster and partially
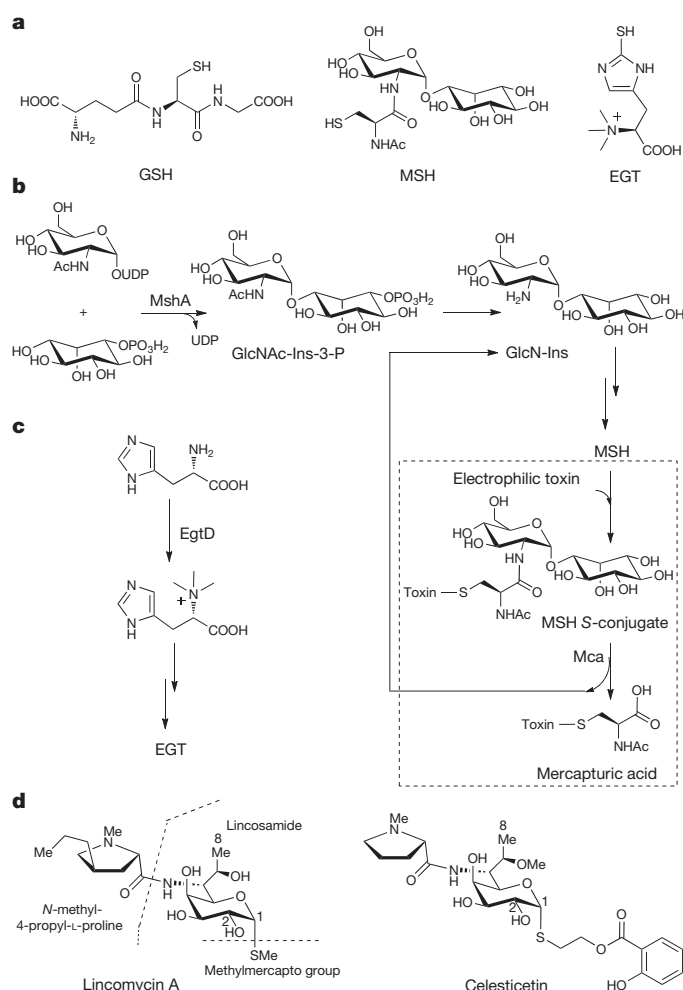
Figure 1 | Representative low-molecular-mass thiols, relevant metabolic pathways, and associated lincosamide natural products. a, Structures of the thiols GSH, MSH and EGT. **b**, Biosynthetic pathway of MSH and its typical associated detoxification process (shown in the dashed rectangle). GlcN-Ins is recyclable as an intermediate or product. **c**, Biosynthetic pathway of EGT. EgtD is an S-adenosyl methionine-dependent protein that catalyses the first reaction. **d**, Structures of the lincosamide antibiotics lincomycin A and celesticetin.

[1]State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China. [2]Huzhou Center of Bio-Synthetic Innovation, 1366 Hongfeng Road, Huzhou 313000, China.
*These authors contributed equally to this work.
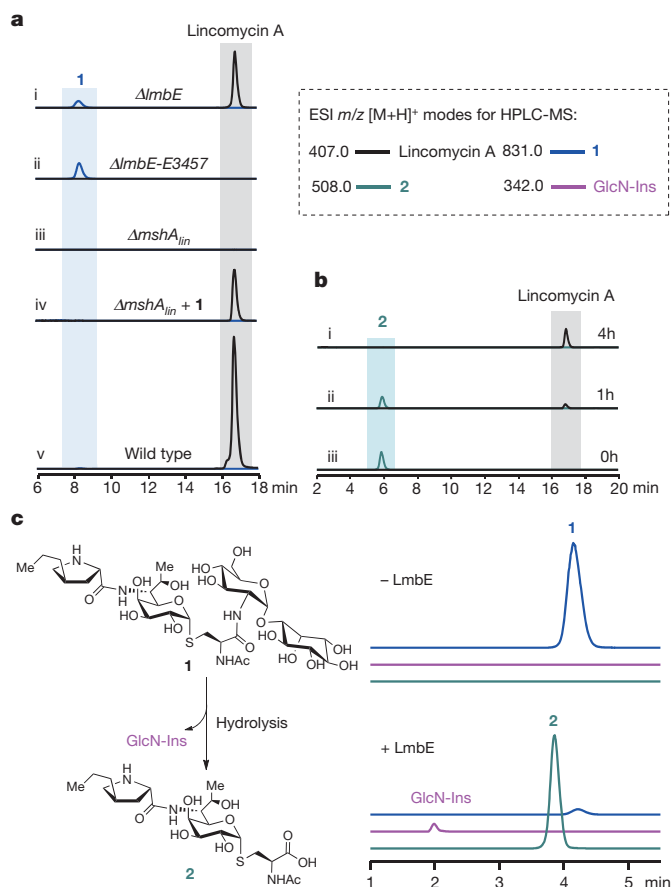
Figure 2 | Characterization of LmbE as a pathway-specific Mca protein to process 1, the MSH *S*-conjugated intermediate, in the biosynthesis of lincomycin A. For high-performance liquid chromatography-mass spectrometry (HPLC-MS) analysis, the electrospray ionization (ESI) *m/z* [M + H]$^+$ modes are indicated in the dashed rectangle. **a**, *In vivo* product profiles of *S. lincolnensis* strains, including the mutants (i, for *ΔlmbE*; ii, for *ΔlmbE-E3457*; iii, for *ΔmshA$_{lin}$*; and iv, for *ΔmshA$_{lin}$* supplemented with ~0.20 mM intermediate **1**) and the wild-type control (v). **b**, Time-dependent biotransformation of mercapturic acid **2** (~0.20 mM) into lincomycin A using the cell homogenate of the *ΔmshA$_{lin}$* mutant strain. **c**, *In vitro* hydrolysis of intermediate **1** to generate **2** and GlcN-Ins (left) in the absence (top right) and presence (lower right) of LmbE.

compensates for the loss of *lmbE*. Sequencing the genome of *S. lincolnensis* revealed three *lmbE* homologues: *lmbE80*, *lmbE447* and *lmbE3457*. The individual inactivation of these genes was performed in the *ΔlmbE* mutant strain, but only the *ΔlmbE-E3457* double mutant strain completely lost lincomycin producibility and had a concomitant increase in **1** (Fig. 2a and Extended Data Fig. 3). Moreover, we expressed and purified LmbE from *Escherichia coli*. This recombinant protein rapidly converted **1** into two products (Fig. 2c): GlcN-Ins and **2**, a mercapturic acid derivative (Supplementary Text). Thus, the involvement of LmbE as a specific Mca protein in lincomycin biosynthesis was established.

MSH biosynthesis begins with the glycosyltransferase (GTase) MshA, which catalyses the formation of 1-*O*-(2-*N*-acetyl)-glucosamine-D-*myo*-inositol-3-phosphate (GlcNAc-Ins-3-P) to afford the essential pseudodisaccharide unit[12] (Fig. 1b). To validate the necessity of MSH for lincomycin biosynthesis, we identified a *mshA* orthologue (Extended Data Fig. 2b), *mshA$_{lin}$*, from the *S. lincolnensis* genome and inactivated it in the wild-type strain. The *ΔmshA$_{lin}$* mutant strain failed to produce MSH along with lincomycin A (Fig. 2a and Extended Data Fig. 4). MSH *S*-conjugate **1** was then added to this mutant, leading to the restoration of lincomycin production (Fig. 2a). Unambiguously, **1** is a key intermediate rather than a detoxified antibiotic in the biosynthetic pathway. Notably, adding ~0.20 mM **1** to the cells yielded ~1.08 mM lincomycin A

after a 5 day cultivation period, representing a ~4-fold increase in product concentration compared with the precursor. This increase could result from the regeneration of intermediate **1** by restoring the MSH cycle, because GlcN-Ins, the product originating from **1** through LmbE-catalysed hydrolysis, is the intermediate in MSH biosynthesis[4,11] (Fig. 1b). The cell homogenate of the *ΔmshA$_{lin}$* mutant strain was capable of transforming ~0.20 mM **2** into a nearly equal amount of lincomycin A (Fig. 2b), further confirming the essentiality of the LmbE-catalysed reaction in the lincomycin biosynthetic pathway, which hydrolyses **1** to generate recyclable GlcN-Ins and intermediate **2**. Processing **2**, including C–S bond cleavage of the cysteinyl group (mechanistically similar to that in the GSH-associated biosynthesis of gliotoxin[17]) and subsequent *S*-methylation of the exposed sulfhydryl group, may eventually afford the sulfur appendage of lincomycin A. Notably, the generation of excretive thiomethyl products has previously been found as an alternative to the mercapturate pathway for xenobiotic detoxification in GSH-mediated metabolism[18].

We next focused on how and when MSH is incorporated to generate *S*-conjugate **1**. Sequence analysis of the *lmb* cluster revealed two closely linked, functionally unassigned genes (Extended Data Fig. 2a), *lmbT* and *lmbV*. *lmbT* encodes a protein belonging to the GTase superfamily. Recently, lincosamide formation has proven to involve the generation of GDP-D-α-D-octose and its associated modifications[19–21]. The phosphonucleotidyl group on the resulting product, GDP-D-α-D-lincosamide (**3**) (Fig. 3), facilitates the attack by nucleophiles such as MSH, probably requiring a GTase activity; however, direct transfer of lincosamide from GDP onto MSH appears unlikely, because GTase-catalysed glycosylation is often envisioned proceeding through S$_N$2 nucleophilic displacement, which would not explain the same α-linkage that is predicted in **3** and found in intermediate **1**. *lmbV* encodes a protein classified into the DinB-2 superfamily (Extended Data Fig. 5), which now includes over ten thousand members with activities presumably related to various low-molecular-mass thiols[22]. These proteins share a conserved DinB-2-like domain combined with variable functional domain(s), indicating that thiols potentially act in different biochemical processes. Notably, a clade phylogenetically relevant to LmbV contains the MSH-maleylpyruvate isomerase (Extended Data Fig. 5), one of the few MSH-dependent proteins that have been biochemically characterized for the isomerization of maleylpyruvate to fumarylpyruate[23,24]. Accordingly, LmbV was proposed to catalyse a MSH-dependent reaction, although the role of MSH remained unknown.

We next established the relevance of LmbT and LmbV to lincomycin biosynthesis, as the inactivation of *lmbT* or *lmbV* completely abolished lincomycin production, which was partially restored by complementing each of the genes in the corresponding mutant strain (Fig. 3a). Surprisingly, a lincomycin analogue, EGT *S*-conjugate **4**, was found in the *ΔlmbV* mutant (Fig. 3 and Supplementary Text). Re-examination of the *ΔmshA$_{lin}$* mutant strain showed a similar product profile in which **4** was the major product (Fig. 3a); evidently, EGT *S*-conjugation is independent of MSH. To correlate this thiol with lincomycin production, we surveyed the *S. lincolnensis* genome and identified the EGT biosynthetic genes[25] (Extended Data Fig. 2c), in which *egtD$_{lin}$*, responsible for triple *N*-methylation of histidine in EGT biosynthesis (Fig. 1c), was chosen for inactivation. The *ΔegtD$_{lin}$* mutant strain produced lincomycin A, the yield of which, however, was significantly reduced at the indicated time points (Fig. 3b). Analysis of the mutant cells revealed a trace amount of EGT (with a yield ~16.8% of that in the wild-type cells, Extended Data Fig. 4), which probably resulted from the exogenous assimilation of the culture medium that included plant-derived, EGT-containing components[4,5]. Over time, the production of lincomycin A from the mutant became much lower than that from the wild-type strain (Fig. 3b), presumably because of the limiting amount of EGT. Altogether, EGT is necessary for lincomycin biosynthesis, whereby its *S*-conjugate **4** is an intermediate instead of a shunt product in the pathway.

Notably, in EGT *S*-conjugate **4**, lincosamide is attached to the thiol via a β-*S*-linkage (Fig. 3 and Supplementary Text). If **1** is the product of
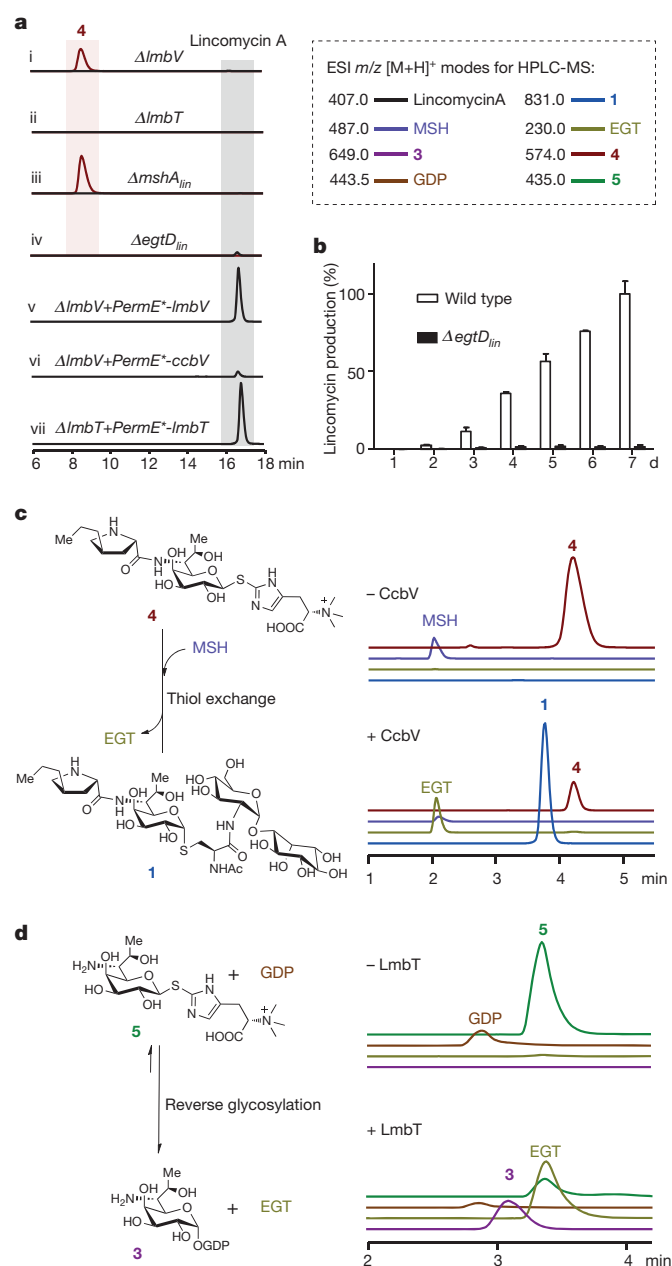
Figure 3 | Characterization of MSH- and EGT-associated biotransformations. For HPLC-MS analysis, the ESI m/z [M + H]$^+$ modes are indicated in the dashed rectangle. **a**, *In vivo* product profiles of *S. lincolnensis* strains, including the mutants $\Delta lmbV$ (i), $\Delta lmbT$ (ii), $\Delta mshA_{lin}$ (iii) and $\Delta egtD_{lin}$ (iv); the $\Delta lmbV$ mutant harbouring homologous $lmbV$ (v) or heterologous $ccbV$ (vi); and the $\Delta lmbT$ mutant harbouring homologous $lmbT$ (vii). PermE* is the constitutive promoter for expressing the erythromycin-resistance gene in *Saccharopolyspora erythraea*. **b**, Time-course analysis of lincomycin production in the wild-type (white) and $\Delta egtD_{lin}$ mutant (black) strains. Mean ± s.e.m. ($n = 3$) are shown. **c**, *In vitro* thiol exchange for converting EGT *S*-conjugate **4** into MSH *S*-conjugate **1** (left) in the absence (top right) and presence of CcbV (lower right). **d**, *In vitro* reverse glycosylation for converting EGT *S*-conjugate **5** into GDP-mediated C8 sugar **3** (left) in the absence (top right) and presence of LmbT (lower right).

LmbV, MSH is probably the agent that performs the nucleophilic attack at C-1 of the lincosamide unit, which is activated by EGT rather than (d)NDP, and the resulting inversion of configuration is consistent with the glycosylation reaction that proceeds via an S$_N$2 displacement to generate the α-*S*-linkage in MSH *S*-conjugate **1**. Therefore, we chemically synthesized MSH (Supplementary Methods), making use of the scaled-up precursor GlcN-Ins, for an *in vitro* assay of LmbV activity.

Despite our numerous attempts, LmbV was highly refractory to various methods of expression. We then selected CcbV, a homologue of LmbV (57% identity) in celesticetin biosynthesis[26], as the catalyst. Celesticetin is a naturally occurring analogue of lincomycin A that possesses the same α-*S*-linkage (Fig. 1d). Heterologous expression of the gene *ccbV* in the $\Delta lmbV$ mutant strain partially restored the production of lincomycin A (Fig. 3a), confirming that CcbV functionally substitutes for LmbV. The *in vitro* conversion of **4** occurred in the presence of CcbV, resulting in the production of **1** along with the release of EGT (Fig. 3c and Extended Data Fig. 6). This reaction is irreversible, as MSH *S*-conjugate **1** was not converted back to EGT *S*-conjugated **4** in the presence of EGT. Therefore, LmbV and its counterpart CcbV form a new type of GTase that catalyses a unique *S*-glycosylation for thiol exchange between EGT and MSH.

We thus considered that the other GTase, LmbT, could transfer lincosamine from GDP onto EGT via S$_N$2 nucleophilic displacement. LmbT was expressed and purified, and to determine its activity, the reverse glycosylation reaction was conducted according to a method established previously using saturated GDP as the nucleophilic agent[27]. Because LmbT did not convert EGT *S*-conjugate **4** to the corresponding GDP-mediated sugar product (Extended Data Fig. 7a), we re-examined the $\Delta lmbT$ mutant strain and found that the PPL moiety had accumulated extensively (Extended Data Fig. 8a). This finding suggested that the *S*-glycosylation of EGT precedes PPL incorporation. Consistently, inactivation of *lmbC*, which encodes an adenylation protein characterized in PPL activation[28], resulted in the production of PPL and an EGT *S*-conjugated lincosamine, **5** (Extended Data Fig. 8a and Supplementary Text). Compared with **4**, **5** lacks the PPL moiety. A similar result was found in the $\Delta lmbN$ and $\Delta lmbD$ mutant *S. lincolnensis* strains (Extended Data Fig. 8a), supporting the corresponding proteins, LmbC, LmbN and LmbD, being responsible for incorporating PPL to transform **5** into **4**. LmbN is a bi-functional protein, and its 1,2-isomerization activity has recently been shown in lincosamine formation[20]. Careful analysis of the protein sequence revealed a peptidyl carrier protein (PCP) domain present at its amino terminus. LmbC is thus considered to activate PPL with ATP and transfer it onto this PCP domain, followed by LmbD-catalysed condensation with EGT *S*-conjugate **5** to afford **4** (Extended Data Fig. 8b). Consequently, LmbC, LmbN (or LmbN-PCP, the N-terminal PCP domain) and CcbD (the homologue in celesticelin biosynthesis[26] that is functionally identical to LmbD) were expressed and purified, and *in vitro* assays showed that these proteins indeed coordinate PPL attachment to generate **4** (Extended Data Fig. 8). Clearly, functionalization of the lincosamine unit by PPL occurs in an EGT *S*-conjugated form, and the PPL-lacking compound **5** probably serves as the product of LmbT-catalysed *S*-glycosylation.

As anticipated, in the presence of LmbT and the co-substrate GDP, **5** was efficiently transformed to GDP-D-α-D-lincosamide **3**, accompanied by EGT release (Fig. 3d and Extended Data Fig. 9). We then validated the forward activity of LmbT to produce **5** and GDP using the substrates **3** and EGT (Extended Data Fig. 7b). The LmbT-catalysed reaction exhibited an equilibrium constant K_{eq} of 1.94 (Extended Data Fig. 9d), indicating that the reverse and forward conversions are comparable; however, the activities of downstream enzymes may drive the pathway forward to produce lincomycin A. Consequently, LmbT represents a new enzyme that employs EGT as a sugar acceptor and catalyses *S*-glycosylation with the naturally rare C8 sugar to generate EGT *S*-conjugate **5**.

The $\Delta lmbC$, $\Delta lmbN$ or $\Delta lmbD$ mutant strain produced a minor product, **6** (Extended Data Fig. 8a), which is a MSH *S*-conjugated lincosamide lacking the PPL moiety (Supplementary Text), suggesting that LmbV tolerated **5** as a substrate for thiol exchange. CcbV, the homologue of LmbV, converted **5** into **6** *in vitro* (Extended Data Fig. 10a); however, the rate of this reaction was much lower than that for the transformation of **4** into **1**. The Mca protein LmbE was not active on **6** and failed to produce mercapturic acid for further processing (Extended Data Fig. 10b). Clearly, lincomycin biosynthesis involves the EGT-mediated
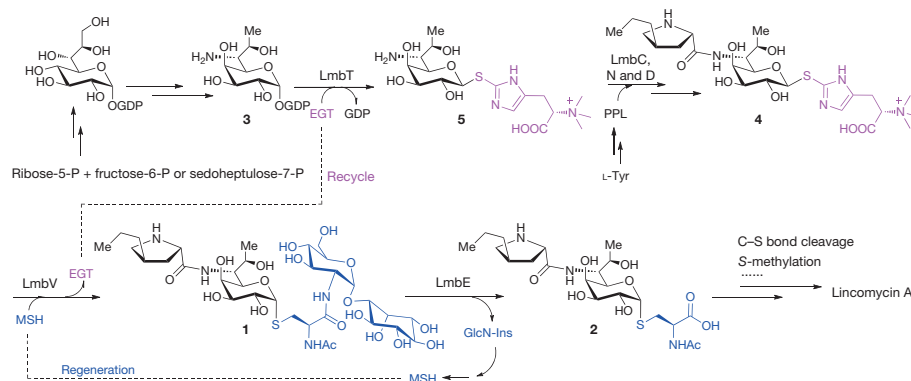
Figure 4 | The EGT (pink) and MSH (blue) programmed biosynthetic pathway of lincomycin A. The recycling of EGT and the regeneration of MSH are shown as dashed lines.

assembly of lincosamide with the building blocks, PPL and MSH, and the MSH-associated post-modifications in a highly ordered process.

We uncovered a constructive role of two different bacterial thiols in *S. lincolnensis* and clarified the biosynthetic pathway of lincomycin A, in which EGT and MSH-associated metabolism creates an extraordinary biochemical strategy for the formation of active molecule (Fig. 4 and Extended Data Fig. 1). Lincomycin biosynthesis involves a convergent pathway to synthesize the L-Tyr-derived PPL moiety and the GDP-activated C8 sugar lincosamide. The thiol EGT serves as a carrier, via the first S-glycosylation (reversible), to channel the lincosamine unit and mediates its condensation with PPL. The thiol MSH goes through the second S-glycosylation (irreversible) to associate with the lincomycin intermediate resulting from EGT and then acts as the sulfur donor for affording the methylmercapto group. Both thiols are recyclable or reproducible, thus maintaining the biosynthetic route to lincomycin A.

The characterization of the low-molecular-mass thiol-programmed biosynthetic pathway largely expands our knowledge regarding the intrinsic, versatile functions of thiols, which are apparently not limited to a protective role against oxidative stress and the neutralization of electrophilic toxins. The lincosamide antibiotic celesticetin probably shares this biosynthetic strategy based on the thiols EGT and MSH, despite their differences in processing the MSH appendage[26]. The involvement of EGT in C8-sugar transfer and activation represents the first biochemical evidence of this thiol in enzymatic reactions and generates interest in nucleotide-independent sugar modifications and associated glycosylations, which have been less appreciated to date[29]. Sulfur is one of the most abundant elements in living organisms and contributes to a large number of biologically active natural products; however, incorporation of this atom has not been well established. Complementing the recent advance of co-opting the sulfur-delivery system of primary metabolism for thiosugar formation[30], we demonstrated that MSH serves as a different source for sulfur incorporation (Fig. 4). This could be a general paradigm because the biosynthetic pathways of several sulfur-containing natural products involve homologues of Mca, the protein responsible for processing the MSH S-conjugate[4,11]. The findings reported here represent a key step towards elucidating the biochemical mechanisms of numerous MSH and EGT-dependent but poorly understood proteins and exploring new features of thiols with regard to their currently unknown associated biochemical processes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Fahey, R. C. Glutathione analogs in prokaryotes. *Biochim. Biophys. Acta* **1830,** 3182–3198 (2013).
2. Deponte, M. Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. *Biochim. Biophys. Acta* **1830,** 3217–3266 (2013).
3. Newton, G. L. *et al.* Bacillithiol is an antioxidant thiol produced in Bacilli. *Nature Chem. Biol.* **5,** 625–627 (2009).
4. Rawat, M. & Av-Gay, Y. Mycothiol-dependent proteins in actinomycetes. *FEMS Microbiol. Rev.* **31,** 278–292 (2007).
5. Cheah, I. K. & Halliwell, B. Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim. Biophys. Acta* **1822,** 784–793 (2012).
6. Mason, D. J., Dietz, A. & DeBoer, C. Lincomycin, a new antibiotic. I. Discovery and biological properties. *Antimicrob. Agents Chemother.* 554–559 (1962).
7. Herr, R. R. & Bergy, M. E. Lincomycin, a new antibiotic. II. Isolation and characterization. *Antimicrob. Agents Chemother.* 560–554 (1962).
8. Lewis, C., Clapp, H. W. & Grady, J. E. *In vitro* and *in vivo* evaluation of lincomycin, a new antibiotic. *Antimicrob. Agents Chemother.* 570–582 (1963).
9. Mason, D. J. & Lewis, C. Biological activity of the lincomycin related antibiotics. *Antimicrob. Agents Chemother.* 7–12 (1964).
10. Newton, G. L., Buchmeier, N. & Fahey, R. C. Biosynthesis and functions of mycothiol, the unique protective thiol of *Actinobacteria. Microbiol. Mol. Biol. Rev.* **72,** 471–494 (2008).
11. Jothivasan, V. K. & Hamilton, C. J. Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Nature Prod. Rep.* **25,** 1091–1117 (2008).
12. Fan, F., Vetting, M. W., Frantom, P. A. & Blanchard, J. S. Structures and mechanisms of the mycothiol biosynthetic enzymes. *Curr. Opin. Chem. Biol.* **13,** 451–459 (2009).
13. Reusser, F. Effect of lincomycin and clindamycin on peptide chain initiation. *Antimicrob. Agents Chemother.* **7,** 32–37 (1975).
14. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Rev. Microbiol.* **12,** 35–48 (2014).
15. Peschke, U., Schmidt, H., Zhang, H.-Z. & Piepersberg, W. Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78–11. *Mol. Microbiol.* **16,** 1137–1156 (1995).
16. Koběrská, M. *et al.* Sequence analysis and heterologous expression of the lincomycin biosynthetic cluster of the type strain *Streptomyces lincolnensis* ATCC 25466. *Folia Microbiol.* **53,** 395–401 (2008).
17. Scharf, D. H. *et al.* Epidithiol formation by an unprecedented twin carbon-sulfur lyase in the gliotoxin. *Angew. Chem. Int. Ed.* **124,** 10211–10215 (2012).
18. Cooper, A. J. L. *et al.* Cysteine S-conjugate β-lyases: important roles in the metabolism of naturally occurring sulfur and selenium-containing compounds, exnobiotics, and anticancer agents. *Amino Acids* **41,** 7–27 (2011).
19. Brahme, N. M. *et al.* Biosynthesis of the lincomycins. 2. Studies using stable isotopes on the biosynthesis of methylthiolincosaminide moiety of lincomycin A. *J. Am. Chem. Soc.* **106,** 7878–7883 (1984).
20. Sasaki, E., Lin, C.-I., Lin, K.-Y. & Liu, H.-w. Construction of the octose 8-phosphate intermediate in lincomycin A biosynthesis: characterization of the reactions catalyzed by LmbR and LmbN. *J. Am. Chem. Soc.* **134,** 17432–17435 (2012).
21. Lin, C.-I., Sasaki, E., Zhong, A. & Liu, H.-w. *In vitro* characterization of LmbK and LmbO: identification of GDP-D-*erythro*-α-D-*gluco*-octose as a key intermediate in lincomycin A biosynthesis. *J. Am. Chem. Soc.* **136,** 906–909 (2014).
22. Newton, G. I., Leung, S. S., Wakabayashi, J. I., Rawat, M. & Fahey, R. C. The DinB superfamily includes novel mycothiol, bacillithiol, and glutathione S-transferases. *Biochemistry* **50,** 10751–10760 (2011).
23. Feng, J. *et al.* The gene *ncgl2918* encodes a novel maleylpyruvate isomerase that needs mycothiol as cofactor and links mycothiol biosynthesis and gentisate assimilation in *Corynebacterium glutamicum. J. Biol. Chem.* **281,** 10778–10785 (2006).
24. Wang, R. *et al.* Crystal structures and site-directed mutagenesis of a mycothiol-dependent enzyme reveal a novel folding and molecular basis for mycothiol-mediated maleylpyruvate isomerization. *J. Biol. Chem.* **282,** 16288–16294 (2007).
25. Seebeck, F. P. *In vitro* reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **132,** 6632–6633 (2010).
26. Čermák, L. *et al.* Hybridization analysis and mapping of the celesticetin gene cluster revealed genes shared with lincomycin biosynthesis. *Folia Microbiol.* **52,** 457–462 (2007).
27. Zhang, C. *et al.* Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313,** 1291–1294 (2006).

28. Kadlčík, S. *et al.* Adaptation of an L-proline adenylation domain to use 4-propyl-L-proline in the evolution of lincosamide biosynthesis. *PLoS ONE* **8,** e84902 (2013).
29. Gantt, R. W., Peltier-Pain, P. & Thorson, J. S. Enzymatic methods for glyco (diversification/randomization) of drugs and small molecules. *Natural Prod. Rep.* **28,** 1811–1853 (2011).
30. Sasaki, E. *et al.* Co-opting sulphur-carrier proteins from primary metabolic pathways for 2-thiosugar biosynthesis. *Nature* **510,** 427–431 (2014).

Supplementary Information is available in the online version of the paper.

METHODS

General materials and methods. Biochemicals and media were purchased from Sinopharm Chemical Reagent Co., Ltd (China), Oxoid Ltd (UK) or Sigma-Aldrich Corporation (USA) unless otherwise stated. Restriction endonucleases were purchased from Thermo Fisher Scientific Co. Ltd (USA). Chemical reagents were purchased from standard commercial sources.

DNA isolation and manipulation in *Escherichia coli* or *Streptomyces* strains were carried out according to standard methods[31,32]. PCR amplifications were carried out on an Applied Biosystems Veriti Thermal Cycler using either Taq DNA polymerase (Vazyme Biotech Co. Ltd, China) for routine genotype verification or Phanta Max Super-Fidelity DNA Polymerase (Vazyme Biotech Co. Ltd, China) for high fidelity amplification. Primer synthesis was performed at Shanghai Sangon Biotech Co. Ltd (China), and DNA sequencing was performed at Shanghai Majorbio Biotech Co. Ltd or Shenzhen BGI in China.

High performance liquid chromatography (HPLC) analysis was carried out on the Agilent 1200 HPLC system (Agilent Technologies Inc., USA). HPLC Electrospray ionization MS (HPLC-ESI-MS) analysis was performed on the Thermo Fisher LTQ Fleet ESI-MS spectrometer (Thermo Fisher Scientific Inc., USA), and the data were analysed using Thermo Xcalibur software. ESI-high resolution MS (ESI-HR-MS) analysis was carried out on the 6230B Accurate-Mass TOF LC/MS System or 6530 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies Inc., USA) and the data were analysed using Agilent MassHunter Qualitative Analysis software. NMR data were recorded on the Bruker DRX400 and Bruker AV500 spectrometers (Bruker Co. Ltd, Germany), or on the Agilent 500 MHz PremiumCompact+ NMR spectrometer (Agilent Technologies Inc., USA).

No statistical methods were used to predetermine sample size.

Gene inactivation and complementation. The inactivation of each gene in *S. lincolnensis* was performed by in-frame deletion to exclude polar effects on downstream gene expression. For complementation *in trans*, the target gene was under the control of *PermE**, the constitutive promoter for expressing the erythromycin-resistance gene in *Saccharopolyspora erythraea*. The genomic DNA of the *S. lincolnensis* wild-type strain served as the template for PCR amplification unless otherwise stated.

For *lmbE* deletion, the 2.52-kb fragment obtained using primers lmbE-L-for and lmbE-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1001. The 2.34-kb fragment obtained using primers lmbE-R-for and lmbE-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1001 to yield the recombinant plasmid pLL1002, in which a 549-bp in-frame coding region of *lmbE* was deleted. To transfer pLL1002 into the lincomycin-producing strain *S. lincolnensis*, conjugation between *E. coli* ET12567-Streptomyces was carried out following the standard procedure[1]. The colonies that were apramycin resistant at 37 °C were identified as integrating mutants, in which a single-crossover homologous recombination event took place. These mutants were cultured for several rounds in the absence of apramycin, and the resulting apramycin-sensitive isolates were subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 1.1-kb product when using primers lmbE-gt-for and lmbE-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1001, in which *lmbE* was in-frame deleted.

For *lmbE*-E80 double deletion, the 1.84-kb fragment obtained using primers lmbE80-L-for and lmbE80-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1003. The 2.06-kb fragment obtained using primers lmbE80-R-for and lmbE80-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1003, yielding the recombinant plasmid pLL1004. Following the procedure described above, pLL1004 was introduced into LL1001 for double-crossover recombination, yielding the recombinant strain LL1002, in which the 447-bp internal fragment of *lmbE80* was deleted in frame. Primers lmbE80-gt-for and lmbE80-gt-rev were used for genotype validation by PCR amplification. Further sequencing of this PCR product confirmed the genotype of LL1002, in which *lmbE80* was also in-frame deleted.

For *lmbE*-E447 double deletion, the 1.81-kb fragment obtained using primers lmbE447-L-for and lmbE447-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1005. The 1.82-kb fragment obtained using primers lmbE447-R-for and lmbE447-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1005, yielding the recombinant plasmid pLL1006. Following the procedure described above, pLL1006 was introduced into LL1001 for double-crossover recombination, yielding the recombinant strain LL1003, in which the 459-bp internal fragment of *lmbE447* was deleted in frame. Primers lmbE447-gt-for and lmbE447-gt-rev were used for genotype validation by PCR amplification. Further sequencing of this PCR product confirmed the genotype of LL1003, in which *lmbE447* was also in-frame deleted.

For *lmbE*-E3457 double deletion, the 1.79-kb fragment obtained using primers lmbE3457-L-for and lmbE3457-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1007. The 1.89-kb fragment

obtained using primers lmbE3457-R-for and lmbE3457-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1007, yielding the recombinant plasmid pLL1008. Following the procedure described above, pLL1008 was introduced into LL1001 for double-crossover recombination, yielding the recombinant strain LL1004, in which the 402-bp internal fragment of *lmbE3457* was deleted in frame. Primers lmbE3457-gt-for and lmbE3457-gt-rev were used for genotype validation by PCR amplification. Further sequencing of this PCR product confirmed the genotype of LL1004, in which *lmbE3457* was also in-frame deleted.

For *mshA_lin* deletion, the 2.25-kb fragment obtained using primers mshA-L-for and mshA-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1009. The 2.18-kb fragment obtained using primers mshA-R-for and mshA-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1009 to yield the recombinant plasmid pLL1010, in which a 633-bp in-frame coding region of *mshA_lin* was deleted. Following the procedure described above, pLL1010 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1005 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 1.09-kb product when using primers mshA-gt-for and mshA-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1005, in which *mshA_lin* was in-frame deleted.

For *lmbT* deletion, the 2.56-kb fragment obtained using primers lmbT-L-for and lmbT-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1011. The 2.56-kb fragment obtained using primers lmbT-R-for and lmbT-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1011 to yield the recombinant plasmid pLL1012, in which a 462-bp in-frame coding region of *lmbT* was deleted. Following the procedure described above, pLL1012 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1006 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 1.04-kb product when using primers lmbT-gt-for and lmbT-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1006, in which *lmbT* was in-frame deleted.

For *lmbV* deletion, the 2.29-kb fragment obtained using primers lmbV-L-for and lmbV-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1013. The 2.30-kb fragment obtained using primers lmbV-R-for and lmbV-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1013 to yield the recombinant plasmid pLL1014, in which a 399-bp in-frame coding region of *lmbV* was deleted. Following the procedure described above, pLL1014 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1007 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 1.23-kb product when using primers lmbV-gt-for and lmbV-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1007, in which *lmbV* was in-frame deleted.

For *egtD_lin* deletion, the 1.98-kb fragment obtained using primers egtD-L-for and edtD-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1019. The 1.75-kb fragment obtained using primers egtD-R-for and egtD-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1019 to yield the recombinant plasmid pLL1020, in which a 465-bp in-frame coding region of *egtD_lin* was deleted. Following the procedure described above, pLL1020 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1010 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 2.05-kb product when using primers egtD-gt-for and egtD-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1010, in which *egtD_lin* was in-frame deleted.

For *lmbC* deletion, the 1.69-kb fragment obtained by using primers lmbC-L-for and lmbC-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1023. The 1.82-kb fragment obtained by using primers lmbC-R-for and lmbC-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1023 to yield the recombinant plasmid pLL1024, in which a 708-bp in-frame coding region of *lmbC* was deleted. Following the procedure described above, pLL1024 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1012 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 0.82-kb product when using primers lmbC-gt-for and lmbC-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1012, in which *lmbC* was in-frame deleted.

For *lmbD* deletion, the 2.54-kb fragment obtained using primers lmbD-L-for and lmbD-L-rev was digested by EcoRI and XbaI, and cloned into the same sites of pKC1139 to yield plasmid pLL1031. The 2.51-kb fragment obtained using primers lmbD-R-for and lmbD-R-rev was digested by XbaI and HindIII and cloned into the same site of pLL1031 to yield the recombinant plasmid pLL1032, in which a

459-bp in-frame coding region of *lmbD* was deleted. Following the procedure described above, pLL1032 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1016 was then subjected to PCR amplification to examine the genotype, as judged by the formation of the desired 0.60-kb product when using primers lmbD-gt-for and lmbD-gt-rev. Further sequencing of this PCR product confirmed the genotype of LL1016, in which *lmbD* was in-frame deleted.

For site-specific mutation of *lmbN*, the 2.01-kb fragment obtained using primers lmbN-L-for and lmbN-L-rev was digested by EcoRI and NehI, and cloned into the same sites of pKC1139 to yield plasmid pLL1033. The 1.94-kb fragment obtained using primers lmbN-R-for and lmbN-R-rev was digested by NheI and HindIII and cloned into the same site of pLL1033 to yield the recombinant plasmid pLL1034, which encoded a S37A mutation that replaced the conserved motif XXXDSL with XXXDAL and had a mutation that replaced CTC (encoding L38) with CTA (resulting no coding change) to introduce the NheI site. Following the procedure described above, pLL1034 was introduced into the *S. lincolnensis* wild-type strain for double-crossover recombination. The resulting strain LL1017 was then subjected to PCR amplification to give a 0.86-kb product (using primers lmbN-gt-for and lmbN-gt-rev). NheI digestion was then carried out on this PCR product to determine the genotype, as judged by the release of the desired 0.11-kb and 0.75-kb fragments. Further sequencing of this PCR product confirmed the genotype of LL1017, in which *lmbN* was site-specifically mutated.

For complementation of *lmbT* in LL1006, the 1.5-kb *lmbT*-containing fragment was amplified by PCR using primers lmbT-C-for and lmbT-C-rev, and then cloned into pMD19-T to yield pLL1015. After digestion with BamHI and XbaI, this 1.5-kb DNA fragment was recovered from pLL1015 and ligated to a 0.45-kb EcoRI/BamHII fragment from pWHM79, and the resulting product was ligated into the EcoRI/XbaI site of pKC1139, yielding the recombinant plasmid pLL1016, in which *lmbT* was under the control of the constitutive promoter *PermE**. pLL1016 was then introduced into LL1006 (*ΔlmbT* mutant) by conjugation, generating the corresponding recombinant strain LL1008 that expressed *lmbT* in trans.

For complementation of *lmbV* in LL1007, a 1.03-kb *lmbV*-containing fragment was amplified by PCR using primers lmbV-C-for and lmbV-C-rev, and then cloned into pMD19-T to yield pLL1017. After digestion with BamHI and XbaI, this 1.03-kb DNA fragment was recovered from pLL1017 and ligated to a 0.45-kb EcoRI/BamHII fragment from pWHM79, and the resulting product was then ligated into the EcoRI/XbaI site of pKC1139, yielding the recombinant plasmid pLL1018, in which *lmbV* was under the control of the constitutive promoter *PermE**. pLL1018 was then introduced into LL1007 (*ΔlmbV* mutant) by conjugation, generating the corresponding recombinant strain LL1009 that expressed *lmbV* in trans.

For heterologous complementation of *ccbV* in LL1007, the 1.15-kb *ccbV*-containing fragment was amplified from the genomic DNA of the celesticetin-producing strain *S. caelestis* NRRL2418 by PCR using primers ccbV-C-for and ccbV-C-rev, and then cloned into pMD19-T to yield pLL1021. After digestion with BglII and XbaI, this 1.03-kb DNA fragment was recovered from pLL1021 and subsequently ligated to a 0.45-kb EcoRI/BamHII fragment from pWHM79; the resulting product was ligated into the EcoRI/XbaI site of pKC1139, yielding the recombinant plasmid pLL1022, in which *ccbV* was under the control of the constitutive promoter *PermE**. pLL1022 was then introduced into LL1007 (*ΔlmbV* mutant) by conjugation, generating the corresponding recombinant strain LL1011 that heterologously expressed *ccbV* in trans.

Production and analysis of lincomycin A, intermediate or shunt product. The *S. lincolnensis* wild-type strain or its derivative was spread on agar plates, which was composed of 19 g of starch, 5 g of soybean meal, 0.5 g of K_2HPO_4, 0.5 g of $MgSO_4·7H_2O$, 1.0 g of KNO_3, 0.5 g of NaCl, 0.01 g of $FeSO_4·7H_2O$ and 20.0 g of agar per litre (pH 7.0~7.5), and then incubated at 28 °C for sporulation and growth. Approximately 1 cm² of the sporulated agar of *S. lincolnensis* was cut, chopped, and inoculated into 25 ml of the seed medium, which was composed of 20 g of starch, 10 g of glucose, 10 g soybean of meal, 30 g of corn steep liquor, 1.5 g of $(NH_4)_2SO_4$ and 5 g of $CaCO_3$ per litre (pH 7.0). After incubation at 28 °C and 220 rpm for 36 h, 5 ml of the seed culture broth was transferred into 50 ml of the fermentation medium, which was composed of 100 g of glucose, 25 g of soybean meal, 2 g of corn steep liquor, 8 g of $(NH_4)_2SO_4$, 0.2 g of KH_2PO_4, 8 g of $NaNO_3$, 5 g of NaCl and 8 g of $CaCO_3$ per litre (pH 7.0). Further incubation was carried out at 28 °C and 220 rpm for 7 days.

For product examination, 500 μl of each fermentation broth was mixed with an equal volume of methanol, and after centrifugation to remove the residue, the supernatant was subjected to HPLC analysis on an Agilent Zorbax column (SB-C18, 5 μm, 4.6 × 250 mm, Agilent Technologies Inc., USA) by isocratic elution of 40% 5 mM ammonium acetate (NH_4Ac) and 60% MeOH for 20 min at a flow rate of 0.6 ml min^{-1}. HPLC-ESI-MS analysis of the supernatant diluted five times with 50% MeOH was carried out on the same column by gradient elution of solvent A (10 mM NH_4Ac) and solvent B (CH_3CN) at a flow rate of 1 ml min^{-1} over a 22-min period as follows: $t = 0$ min, 10% B; $t = 9$ min, 10% B; $t = 17$ min, 60% B; $t = 18$ min, 60% B; $t = 20$ min, 10% B; and $t = 22$ min, 10% B (mAU at 210 nM). The data were analysed using Thermo Xcalibur software. The concentration of lincomycin A was estimated by HPLC using the commercially available product as the standard.

For biotransformation of **2** (the mercapturic acid derivative) into loncomycin A by the cell homogenate of the *ΔmshA$_{lin}$* mutant strain, LL1005, the mycelia from 50 ml of the 6-day fermentation culture broth was collected, washed twice, and then re-suspended in 10 ml of 50 mM Tris-HCl buffer (pH 7.5). Sonication for 20 min on ice, followed by centrifugation at 4 °C to remove the cellular debris, resulted in the supernatant that was used for the conversion of **2**. Each biotransformation was conducted in 100 μl of the mixture, which contained 1 μl of **2** (giving a final concentration of 0.20 mM) and 99 μl of the supernatant of LL1005, and incubated at 30 °C for 0, 1 or 4 h. After quenching the reaction by addition of an equal volume of methanol, 20 μl of each biotransformation mixture was centrifuged and subjected to HPLC-ESI-MS analysis of **2** consumption and of lincomycin A production.

For feeding of the MSH S-conjugate **1** to LL1005 (*ΔmshA$_{lin}$*), **1** was added into the culture broth of LL1005 on the 4th day of fermentation (giving a final concentration of approximately 0.20 mM). In a further 5-day incubation period, the production of lincomycin A was monitored daily by HPLC and HPLC-ESI-MS as described above.

Production and analysis of MSH and EGT in *S. lincolnensis*. Approximately 1 cm² of the sporulated agar of *S. lincolnensis* or its derivative was cut, chopped, and inoculated into 25 ml of YEME medium and then incubated at 28 °C and 220 rpm for 36 h. 5 ml of the resulting culture broth was added into 100 ml of the same medium for scale-up and then further incubated at 28 °C and 220 rpm for 3 days. The mycelia were then harvested by centrifugation at 4 °C and 5,000 rpm for 15 min.

For thiol extraction, derivatization and detection, the procedure described by Fahey and Newton[33] was used with slight modifications. Approximately 200 mg (wet weight) of the freshly harvested mycelia was weighed in a 5 ml microcentrifuge tube and then re-suspended in 2 ml of a mixture of 50% CH_3CN and 50% 2 mM monobromobimane (mBBr) dissolved in 20 mM Tris-HCl (pH 8.0) buffer for sonication. The mixture was incubated in the dark at 60 °C for 15 min, acidified with 5 μl of methanesulfonic acid (5N) and then centrifuged at 12,000 rpm for 10 min to remove the debris before storing at −80 °C. HPLC-ESI-HR-MS analysis of the resulting thiol-mBBr derivatives was carried out on a Agilent Zorbax column (Extend-C18, 1.8 μm, 2.1 × 50 mm, Agilent Technologies Inc., USA) by gradient elution of solvent A (H_2O) and solvent B (CH_3CN) at a flow rate of 0.2 ml min^{-1} over a 15 min period as follows: $t = 0$ min, 5% B; $t = 5$ min, 5% B; $t = 10$ min, 50% B; $t = 12$ min, 50% B; $t = 13$ min, 5% B; and $t = 15$ min, 5% B (mAU at 370 nm).

Protein expression and purification. The recombinant proteins LmbE, CcbV, LmbT, LmbC, LmbN, LmbN-PCP and CcbD, all of which were in an N-terminal 6× His-tagged form, were overproduced in *E. coli* BL21(DE3) and purified by Ni-affinity followed by desalting.

The genes of recombinant proteins LmbE, CcbV, LmbT, LmbC, LmbN, LmbN-PCP and CcbD were PCR amplified from *S. lincolnensis* or *S. caelestis* genomic DNA using primers with engineered NedI and HindIII restriction sites. The PCR-amplified gene fragments were purified, digested with NdeI and HindIII and ligated into a pET28a(+) vector (Novagen) that had been digested with the same enzymes. The resultant plasmids were used to transform *E. coli* BL21(DE3) for protein over-expression. The *E. coli* transformant cultures were incubated in Luria-Bertani (LB) medium containing 50 μg ml^{-1} kanamycin at 37 °C and 250 rpm until the cell density reached 0.6–0.8 at OD$_{600\,nm}$. To induce protein expression, IPTG (0.1 mM) was added to the cultures, which were further incubated at 16 °C for 40–48 h. The cells were harvested by centrifugation and stored at −80 °C before lysis. The thawed cells were re-suspended in lysis buffer containing 50 mM Tris-HCl (pH 7.6), 300 mM NaCl, 10 mM imidazole and 10% (v/v) glycerol. After disruption by a low-temperature ultra-high-pressure cell disrupter (FB-110X, Shanghai Litu Mechanical Equipment Engineering Co., Ltd, China or JN-02HC, JNBIO, China), the soluble fraction was collected, subjected to purification of each target protein by using a HisTrap FF column (GE Healthcare, USA) and then desalted using a PD-10 Desalting Column (GE Healthcare, USA) according to the manufacturers' protocols. The resulting proteins were concentrated and stored at −80 °C for *in vitro* assays. The purity of the proteins was examined by 12% SDS–PAGE analysis, and the concentration was determined by Bradford assay using bovine serum albumin (BSA) as the standard. For LmbN-PCP, HPLC-HR-ESI-Q-TOF MS (Agilent Technologies Inc., USA) analysis indicated that the recombinant protein purified from *E. coli* BL21(DE3) was fully phosphopantetheinylated into a holo form (m/z [M + H]$^+$ calculated 11,842.94, found 11,843.42).

Determination of the amidase activity of LmbE *in vitro*. The assays were carried out at 30 °C for 5 h in a 50 μl reaction mixture containing 50 mM Tris-HCl (pH 7.5) and 2 mM substrate (MSH S-conjugate) in the presence of 2 μM (for **1**) or 20 μM (for **6**) LmbE. Reactions in the absence of the enzyme were used as negative controls.

An equal volume of methanol was added into each mixture to terminate the reaction. After removal of the denatured protein by centrifugation, the reaction mixtures were subjected to HPLC-ESI-MS analysis on a Phenomenex Luna C18(2) column (5 μm, 4.6 × 250 mm, USA) by isocratic elution of 90% 10 mM NH$_4$Ac and 10% CH$_3$CN for 15 min period at a flow rate of 1 ml min^{-1}.

Characterization of the CcbV-catalysed reaction *in vitro*. The assays were carried out at 30 °C for 2 h in a 50 μl reaction mixture. For substrate **4**, the mixture contained 50 mM Tris-HCl (pH 8.0), 1 mM MSH, 2 mM TCEP, 1 mM EGT *S*-conjugate (**4**) and 40 μM CcbV. For substrate **5**, the mixture contained 50 mM Tris-HCl (pH 8.0), 2 mM MSH, 2 mM TCEP, 2 mM EGT *S*-conjugate (**5**) and 100 μM CcbV. Reactions in the absence of the enzyme were used as negative controls. The termination of each reaction and analysis of the resulting MSH *S*-conjugate and EGT were carried out according to the methods described above for LmbE-catalysed conversion. HPLC analysis of EGT production was performed on a COSMOSIL HILIC Packed Column (5 μm, 4.6 × 250 mm, Nacalai Tesque Inc., Japan) by isocratic elution of 30% 10 mM NH$_4$Ac and 70% CH$_3$CN for 20 min at a flow rate of 1 ml min^{-1} (mAU at 260 nM). The commercially available EGT (Enzo Life Sciences Inc., USA) was used as the standard.

To evaluate the pH dependence, each reaction was performed in triplicate at 30 °C for 1 h in a 25 μl reaction mixture containing 1 mM MSH, 2 mM TCEP, 1 mM **4** and 20 μM CcbV in 50 mM PIPES (pH 6.0–7.0) or 50 mM Tris-HCl (pH 7.5–9.0) buffer.

For the kinetic analysis, a time course was carried out to determine the initial rate conditions in a 25 μl of reaction mixture containing 50 mM Tris-HCl buffer (pH 8.0), 1 mM MSH, 1 mM TCEP, 1 mM **4** and 20 μM CcbV. The reactions were initiated by the addition of CcbV, incubated at 30 °C, and then terminated by 25 μl of methanol at 3, 5, 7, 10, 16, 25, 45 and 60 min. The samples were subjected to the same work-ups and HPLC analysis of EGT production as described above. The production of EGT, linear with respect to time during 0–20 min, was fitted into a linear equation to obtain the initial velocity. To determine the kinetic parameters for substrate **4**, the reactions were carried out at 30 °C for 20 min, each in 25 μl of mixture containing 50 mM Tris-HCl buffer (pH 8.0), 2 mM MSH, 2 mM TCEP, and 20 μM CcbV, and varying the concentration of substrate **4** (0.02, 0.05, 0.10, 0.20, 0.50, 1.00 and 2.00 mM). To determine the kinetic parameters for the substrate MSH, the reactions were carried out at 30 °C for 20 min, each in 25 μl of mixture containing 50 mM Tris-HCl buffer (pH 8.0), 5 mM **4**, 2 mM TCEP, and 20 μM CcbV, and varying the concentration of substrate MSH (0.02, 0.05, 0.10, 0.20, 0.50 and 1.00 mM). All assays were performed in triplicate, and each conversion was analysed by HPLC for EGT production as described above. The resulting initial velocities were then fitted to the Michaelis–Menten equation using GraphPad Prism5 software (GraphPad Software, Inc., USA) to extract the K_m and k_{cat} parameters.

Characterization of the LmbT-catalysed reaction *in vitro*. The reverse glycosylation assays of LmbT were carried out at 30 °C for 2 h in a 50 μl reaction mixture. For substrate **5**, the mixture contained 50 mM Tris-HCl (pH 7.5), 2 mM MgCl$_2$, 1 mM **5**, 2 mM GDP and 4 μM LmbT. For substrate **4**, the mixture contained 50 mM Tris-HCl (pH 7.5), 2 mM MgCl$_2$, 2 mM **4**, 4 mM GDP and 20 μM LmbT. The forward glycosylation assays were carried out at 30 °C for 2 h in a 50 μl reaction mixture containing 50 mM Tris-HCl (pH 7.5), 2 mM MgCl$_2$, 1 mM **3**, 1 mM EGT and 4 μM LmbT. Reactions in the absence of the enzyme were used as negative controls. The termination of each reaction was carried out according to the methods described above for LmbE-catalysed conversion. The reaction mixtures containing **5** were subjected to HPLC-ESI-MS analysis on a Phenomenex Luna C18(2) column by isocratic elution of 95% 10 mM NH$_4$Ac and 5% CH$_3$CN for 15 min at a flow rate of 1 ml min^{-1}.

To evaluate the pH dependence, each reaction was performed in triplicate at 30 °C for 1 h in 25 μl of reaction mixture containing 2 mM MgCl$_2$, 4 mM GDP, 2 mM **5** and 4 μM LmbT in 50 mM PIPES (pH 6.0–7.0) or 50 mM Tris-HCl (pH 7.5–9.0) buffer.

For the kinetic analysis, a time course was carried out to determine the initial rate conditions in 25 μl of reaction mixture containing 50 mM Tris-HCl (pH 7.5), 2 mM MgCl$_2$, 1 mM **5** and 2 mM GDP and 4 μM LmbT. The reactions were initiated by the addition of LmbT, incubated at 30 °C, and then terminated by 25 μl of methanol at 3, 5, 7, 10, 16, 25, 45 and 60 min. The samples were subjected to the same work-ups and HPLC analysis of EGT production as described above. The production of EGT, linear with respect to time during 0–20 min, was fitted into a linear equation to obtain the initial velocity. To determine the kinetic parameters for substrate **5**, the reactions were carried out at 30 °C for 20 min, each in 25 μl of mixture containing 50 mM Tris-HCl buffer (pH 7.5), 2 mM MgCl$_2$, 2 mM GDP, and 4 μM LmbT, and varying the concentration of substrate **5** (0.02, 0.05, 0.10, 0.20, 0.50, 1.00, 2.00 and 4.00 mM). To determine the kinetic parameters for substrate GDP, the reactions were carried out at 30 °C for 20 min, each in 25 μλ of mixture containing 50 mM Tris-HCl buffer (pH 7.5), 2 mM MgCl$_2$, 4 mM **5**, and 4 μM LmbT, and varying the concentration of substrate GDP (0.02, 0.05, 0.10, 0.20, 0.50, 1.00

and 2.00 mM). All assays were performed in triplicate, and each conversion was analysed by HPLC for EGT production as described above. The resulting initial velocities were then fitted to the Michaelis–Menten equation using GraphPad Prism5 software (GraphPad Software, Inc., USA) to extract the K_m and k_{cat} parameters.

To determine the equilibrium constant (K_{eq}) of the LmbT-catalysed reaction, the experiment was performed according to the method described previously[34]. K_{eq} was measured by performing a series of saturated reactions, in which the concentration ratio of [GDP]/[**3**] varied from 1/3 to 5 on the premise that the ratio of [**5**]/[EGT] was fixed at 1. The total concentrations of [**3**] + [GDP] and [**5**] + [EGT] were both kept at 4 mM. The reaction was performed in a 25 μl mixture containing 50 mM Tris-HCl (pH 7.5), 2 mM MgCl$_2$ and 4 μM LmbT at 30 °C for 3 h to reach equilibrium. The change of [EGT] was monitored by HPLC as described previously and plotted against the ratio of [GDP]/[**3**]. The equilibrium constant was subsequently determined from the equation $K_{eq} = ([GDP]/[3]) \times ([5]/[EGT])$.

Characterization of PPL incorporation *in vitro*. To convert holo-LmbN-PCP into PPL-acylated LmbN-PCP, the reaction was carried out at 30 °C for 3 h in a 50 μl reaction mixture containing 50 mM Tris-HCl (pH 7.5), 2 μM LmbC, 2 mM PPL, 5 mM ATP, 10 mM MgCl$_2$, 2 mM TCEP, and 1 mM CoA in the presence of 100 μM LmbN-PCP. Reactions in the absence of LmbC or ATP were used as negative controls. For product examination, the reaction mixtures were subjected to HPLC-HR-ESI-Q-TOF MS (Agilent Technologies Inc., USA) analysis on an Agilent Zorbax column by gradient elution of solvent A (H$_2$O containing 0.1% formic acid) and solvent B (CH$_3$CN containing 0.1% formic acid) at a flow rate of 0.2 ml min^{-1} over a 45-min period as follows: $t = 0$ min, 10% B; and $t = 45$ min, 95% B (mAU at 210 nm).

For PPL incorporation, the assays were carried out at 30 °C for 3 h in a 50 μl reaction mixture containing 50 mM Tris-HCl (pH 7.5), 2 μM LmbC, 2 mM PPL, 5 mM ATP, 10 mM MgCl$_2$, 2 mM TCEP, 1 mM CoA, and 2 mM substrate **5** (MSH *S*-conjugate) in the presence of 50 μM LmbN or LmbN-PCP and 10 μM CcbD. Reactions in the absence of each enzyme were used as negative controls. The termination of each reaction and HPLC-ESI-MS analysis was carried out according to the methods described above for LmbE-catalysed conversion.

Compound isolation and purification. For compounds **1** and **4** (from the *ΔlmbE* and *ΔlmbV* mutant strain, respectively), 100 g of Amberlite XAD-2 resin (Rohm and Haas Co., USA) was incubated with 1 l of each fermentation culture broth overnight to absorb the target compound. After filtration, the resin was backwashed with 2 l of water, and then eluted with 3 l of MeOH. The eluent was evaporated *in vacuo* to a crude extract. The resultant residue was re-dissolved in 5 ml of water, and then loaded onto a Sephadex LH20 column (3.5 × 200 cm, GE Healthcare, USA) by eluting 500 ml of MeOH at a flow rate of 0.5 ml min^{-1}. According to ESI-MS analysis, the fractions containing the target compound were combined, evaporated *in vacuo* and then loaded onto an Agilent Zorbax column (SB-C18, 5 μm, 9.4 × 250 mm, Agilent Technologies Inc., USA) by isocratic elution of 40% 5 mM NH$_4$Ac and 60% MeOH for 12 min at a flow rate of 2 ml min^{-1} (mAU 210 nm). After a similar work-up for fractionation and concentration, further purification was carried out on a COSMOSIL HILIC Packed Column (5 μm, 10 × 250 mm, Nacalai Tesque Inc., Japan) by isocratic elution of 21% 10 mM NH$_4$Ac and 79% CH$_3$CN for 70 min at a flow rate of 2 ml min^{-1} (mAU 210 nm).

For PPL (from the *ΔlmbT* mutant strain), after a similar procedure with Amberlite XAD-2 resin, further HPLC semi-preparation purification was carried out twice on an Agilent Zorbax column by isocratic elution of 20% 5 mM NH$_4$Ac and 80% MeOH for 30 min at a flow rate of 2 ml min^{-1} (mAU at 210 nm).

For compounds **5** and **6** (from the *ΔlmbC* mutant strain), 100 g of Amberlite XAD-16 resin (Rohm and Haas Co., USA) was incubated with 1 l of fermentation culture broth overnight to remove most of the impurity. After filtration and concentration, the resultant residue was then loaded onto a Sephadex LH20 column by eluting 500 ml of H$_2$O:MeOH (1:1) at a flow rate of 0.5 ml min^{-1}. According to ESI-MS analysis, the fractions containing compound **5** or **6** were combined, evaporated *in vacuo* and then loaded onto a COSMOSIL HILIC Packed Column by isocratic elution of 30% 10 mM NH$_4$Ac and 70% CH$_3$CN for 30 min at a flow rate of 2 ml min^{-1} (mAU at 210 nm). For compound **6**, further purification was carried out on a Sephadex G10 column (1.5 × 120 cm, GE Healthcare, USA) by eluting H$_2$O at a flow rate of 0.2 ml min^{-1}.

For compounds **2** and GlcN-Ins (from the LmbE-catalysed reaction), the assay was scaled up and carried out at 30 °C for 2 h in 20 ml of mixture containing 200 μM LmbE, 1 mM pure compound **1** and 50 mM Tris-HCl buffer (pH 7.5). After filtration with an ultra-filtration membrane (Amicon YM-30, Millipore) to remove protein, the solution was loaded onto an Agilent Zorbax column by gradient elution of solvent A (H$_2$O) and solvent B (CH$_3$CN) at a flow rate of 2 ml min^{-1} over a 18-min period as follows: $t = 0$ min, 5% B; $t = 5$ min, 5% B; $t = 10$ min, 55% B; $t = 16$ min, 55% B; and $t = 18$ min, 5% B (mAU at 210 nm). The fraction containing GlcN-Ins was concentrated and then loaded onto a COSMOSIL HILIC Packed Column by

isocratic elution of 25% 10 mM NH_4Ac and 75% CH_3CN for 40 min at a flow rate of 2 ml min^{-1}. GlcN-Ins was examined using a refractive index (RI) detector. For compound **2**, further purification was carried out on the same column by the same isocratic elution for 60 min (mAU at 210 nm).

For compound **3** (from the LmbT-catalysed reverse glycosylation reaction), the assay was scaled up and carried out at 30 °C for 5 h in 10 ml of mixture containing 4 μM LmbT, 2 mM pure compound **5**, 2 mM GDP, 1 mM $MgCl_2$ and 50 mM Tris-HCl buffer (pH 7.5). After filtration with an ultra-filtration membrane to remove protein, the solution was concentrated and then subjected to a Sephadex G10 column by eluting H_2O at a flow rate of 0.1 ml min^{-1}.

Chemical synthesis of MSH. The synthesis was carried out according to the procedures described previously[35,36] with slight modifications. For details, see Supplementary Methods.

31. Kieser, T. *et al. Practical Streptomyces Genetics* (John Innes Foundation, 2000).
32. Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* 4th edn (Cold Spring Harbor Laboratory Press, 2012).
33. Fahey, R. C. & Newton, G. L. Determination of low molecular weight thiols using monobromobimane fluorescent labeling and high-performance liquid chromatography. *Methods Enzymol.* **143,** 85–96 (1987).
34. Zhang, C. *et al.* Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313,** 1291–1294 (2006).
35. Lee, S. & Rossaza, J. P. First total synthesis of mycothiol and mycothiol disulfide. *Org. Lett.* **6,** 365–368 (2004).
36. Ajayi, K., Thakur, V. V., Lapo, R. C. & Knapp, S. Intramolecular α-glucosaminidation: synthesis of mycothiol. *Org. Lett.* **12,** 2630–2633 (2010).
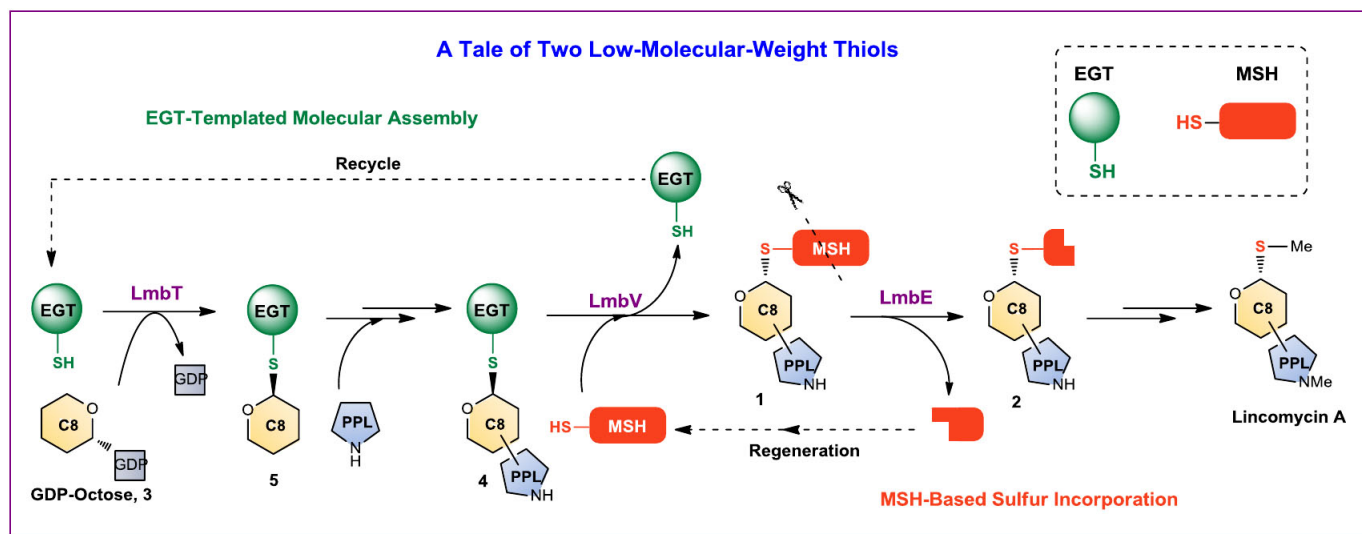
Extended Data Figure 1 │ Constructive role of MSH (orange) and EGT (green) in lincomycin biosynthesis (shown as cartoon models).
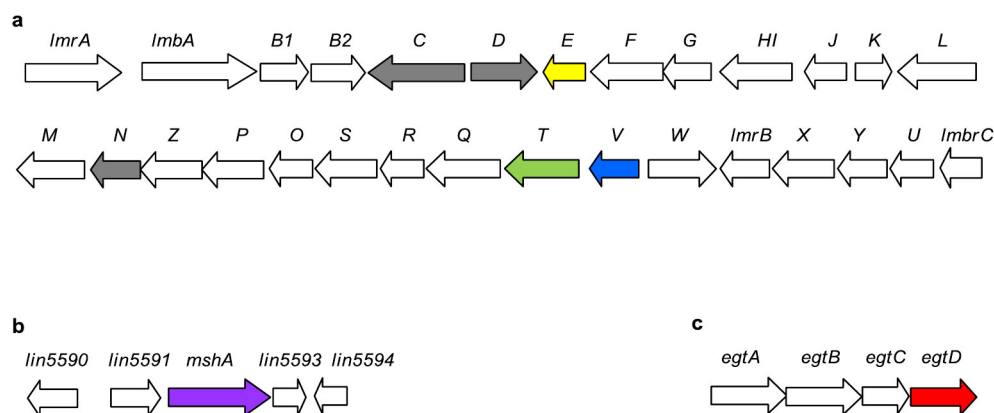
Extended Data Figure 2 | Genes and/or clusters relevant to lincomycin biosynthesis in *S. lincolnensis*. a, Biosynthetic gene cluster of lincomycin A. The *mca* homologue *lmbE* and two GTase-encoding genes, *lmbV* and *lmbT*, are shown in yellow, blue and green, respectively. The genes responsible for PPL incorporation (*lmbC*, *lmbD* and *lmbN*) are shown in grey. **b**, The location of *mshA*$_{lin}$ (shown in purple), which is not clustered with the other genes responsible for MSH biosynthesis, in the genome of *S. lincolnensis*. The flanking genes *lin5590*, *lin5591*, *lin5593* and *lin5594* share sequence homology with genes encoding the methylase (WP_026151264.1) from *S. prunicolor*, the DUF899-like protein (WP_004004056.1) from *S. viridochromogenes*, the YbjN-like protein (WP_020130332.1) from *Streptomyces* sp. 303MFCol5.2 and the hypothetical protein (WP_004000160.1) from *S. viridochromogenes*, respectively. **c**, The biosynthetic gene cluster of EGT. The gene *egtD*$_{lin}$ is shown in red.
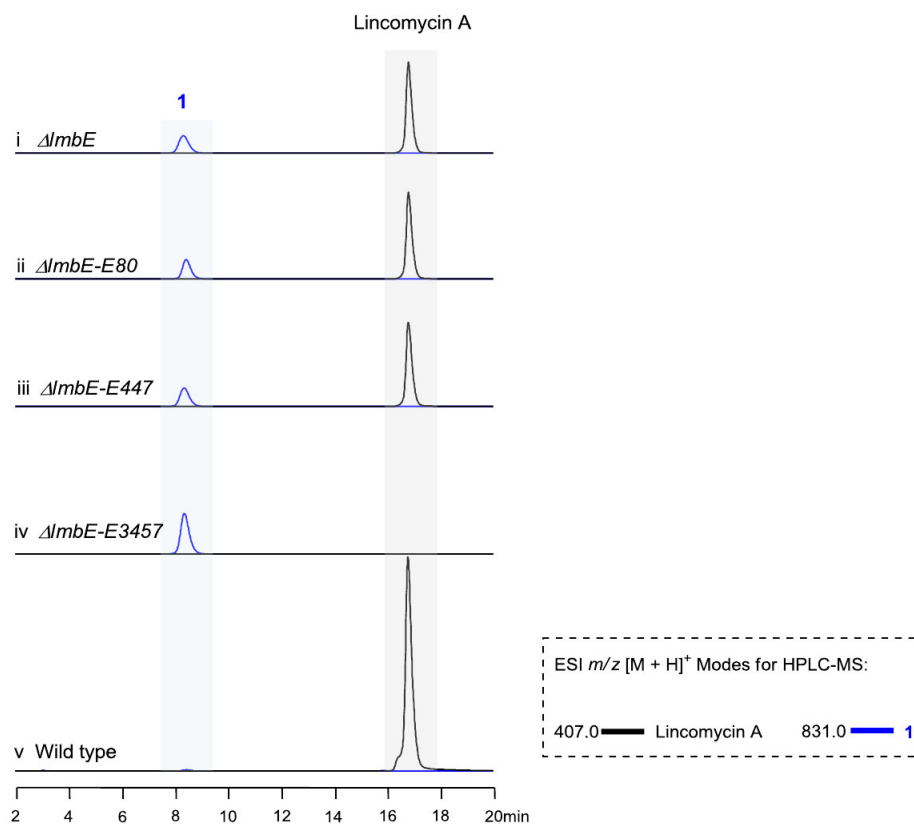
Extended Data Figure 3 | The production of MSH S-conjugate 1 and lincomycin A in various *S. lincolnensis* strains. The strains include the mutants (i, for *ΔlmbE*; ii, for *ΔlmbE-E80*; iii, for *ΔlmbE-E447*; and iv, for *ΔlmbE-E3457*) and the wild-type control (v). For HPLC-MS analysis, the ESI m/z [M + H]$^+$ modes are indicated in the dashed rectangle.
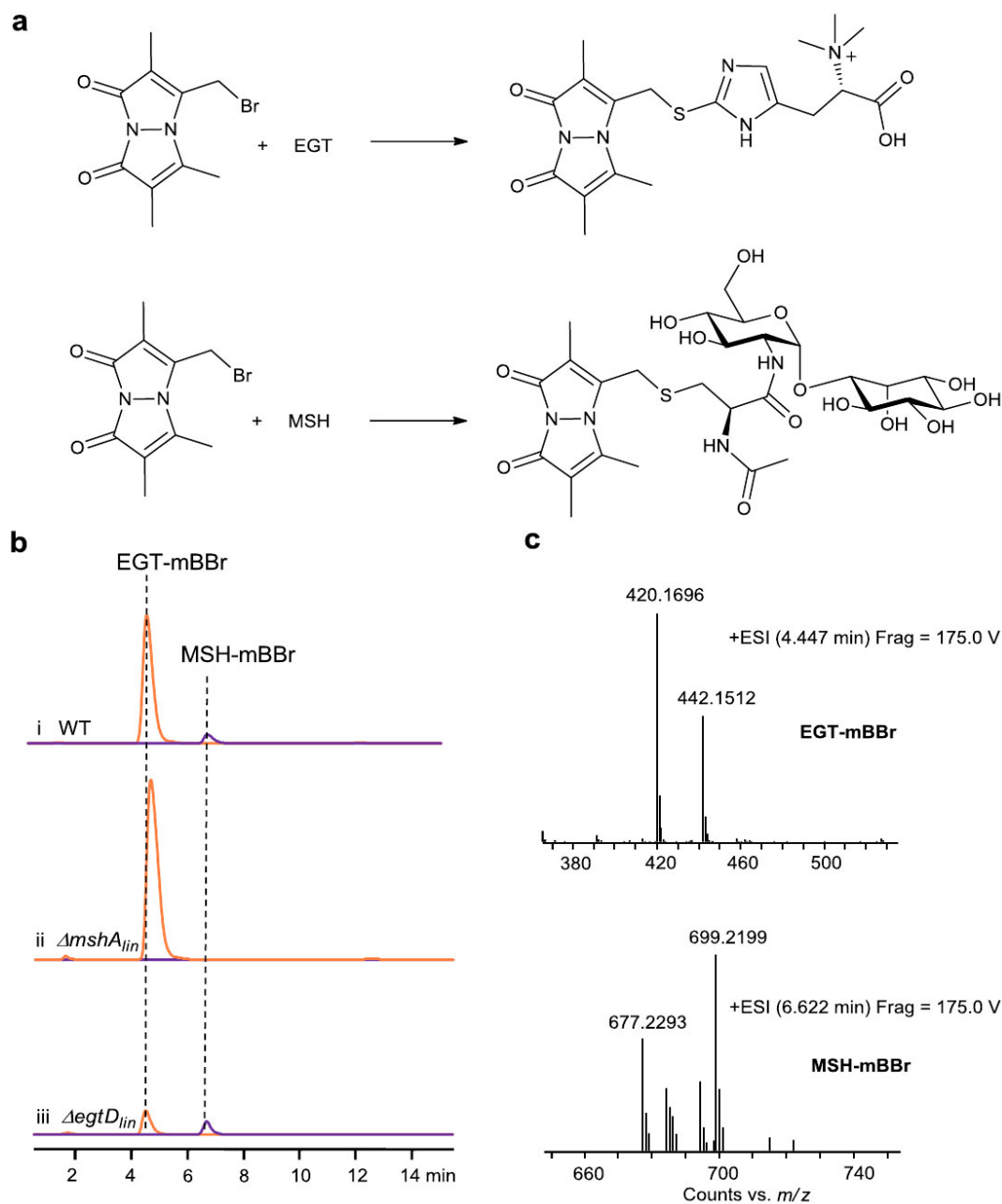
Extended Data Figure 4 | Analysis of MSH and EGT production in the **_ΔmshA__{_lin_} and _ΔegtD__{_lin_} mutant _S. lincolnensis_ strains. a**, Derivatization of MSH and EGT with mBBr to generate the corresponding *S*-conjugates. **b**, HPLC-HR-MS analysis of MSH-mBBr (calculated for $C_{27}H_{41}N_4O_{14}S^+$ $[M + H]^+$ 677.2340) and EGT-mBBr (calculated for $C_{19}H_{26}N_5O_4S^+$ $[M + H]^+$ 420.1706) in the wild-type control (i), LL1005 (ii, *ΔmshA*_{*lin*} mutant) and LL1010 (iii, *ΔegtD*_{*lin*} mutant). **c**, The HR-ESI-MS spectra of EGT-mBBr (top) and MSH-mBBr (below).
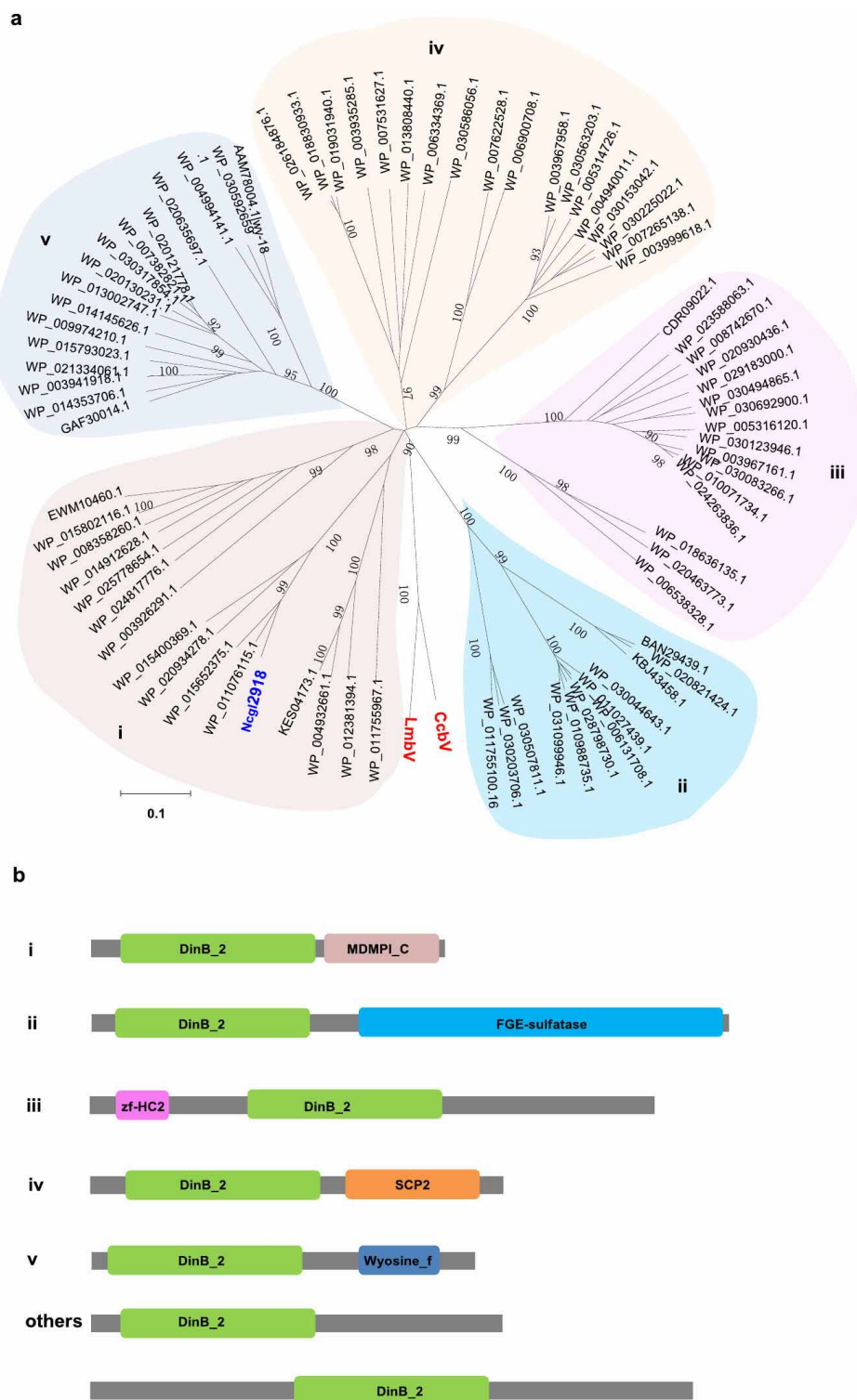
Extended Data Figure 5 | Phylogenetic analysis in the DinB-2 superfamily.
a, LmbV (from *S. lincolnensis*) and CcbV (from *S. caelestis*), shown in red, with selected DinB-2-like proteins (from various actinomycetes, in which the thiol MSH is dominant) in the phylogenetic tree. The evolutionary distances were computed using the p-distance method. The support for grouping clades i, ii, iii, iv and v (shaded in different colours) is indicated by bootstrap values. The known MSH-maleylpyruvate isomerase Ncgl2918 is shown in blue. **b**, Typical domain organization of the DinB-2-like proteins. The conserved DinB-2 domain is shown in green. Clade i features the C-terminal MDMPI-C domain. Proteins containing this domain include the MSH-maleylpyruvate isomerases, such as Ncgl2918. Clade ii features the C-terminal FGE-sulfatase domain,

which is found in eukaryotic proteins required for post-translational modification to produce sulfatases, which are essential for the degradation and remodelling of sulfate esters. Clade iii features the N-terminal zf-HC2 domain, which contains a putative zinc-finger binding motif and is found in some anti-sigma factor proteins. Clade iv features the C-terminal SCP2 domain involved in binding sterols. Clade v features the C-terminal wyosine_f domain. Some proteins containing this domain appear to be important in wyosine base formation in a subset of phenylalanine-specific tRNAs. 'Others' indicate a number of DinB-2-like proteins that possess an unknown domain(s) either at the C terminus or at both the C and N termini of the proteins.
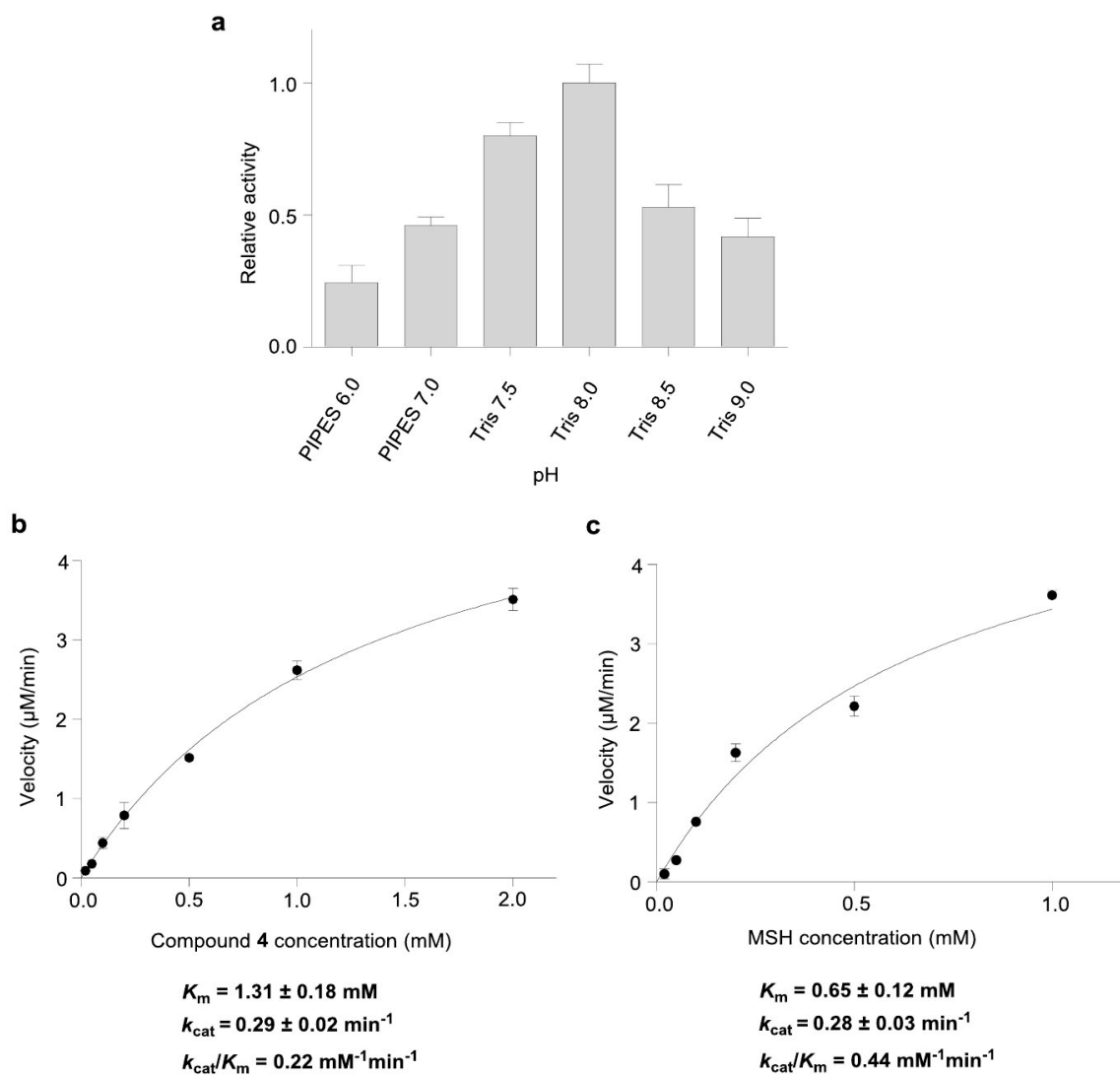
Extended Data Figure 6 | Kinetic analysis of CcbV-catalysed thiol exchange. a, pH dependence. The activity of CcbV in 50 mM PIPES (pH 6.0-7.0) or 50 mM Tris-HCl (pH 7.5–9.0) buffer was measured. **b, c**, Determination of the steady-state kinetic parameters for substrate **4** and for MSH, respectively. The error bars are standard error of mean ($n = 3$).
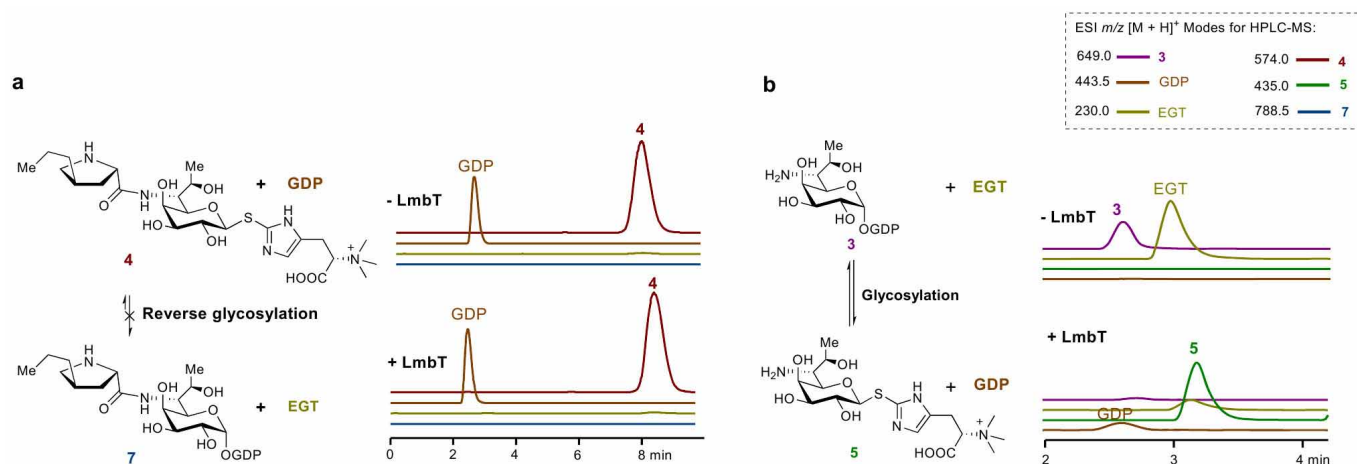
Extended Data Figure 7 | Characterization of LmbT-catalysed reverse and forward glycosylation. For HPLC-MS analysis, the ESI m/z $[M + H]^+$ modes are indicated in dashed rectangle. **a**, Examination of the acylated C8-sugar transfer in the presence of LmbT, which showed that LmbT was unable to utilize **4** as a substrate for reverse glycosylation to generate the predicted GDP-D-α-D-sugar **7** (left) in the absence (top right) and in the presence (lower right) of LmbT. **b**, Characterization of LmbT-catalysed forward glycosylation. LmbT used **3** as a substrate for glycosylation to generate the GDP-D-α-D-sugar **5** (left) in the absence (top right) and in the presence (lower right) of LmbT.
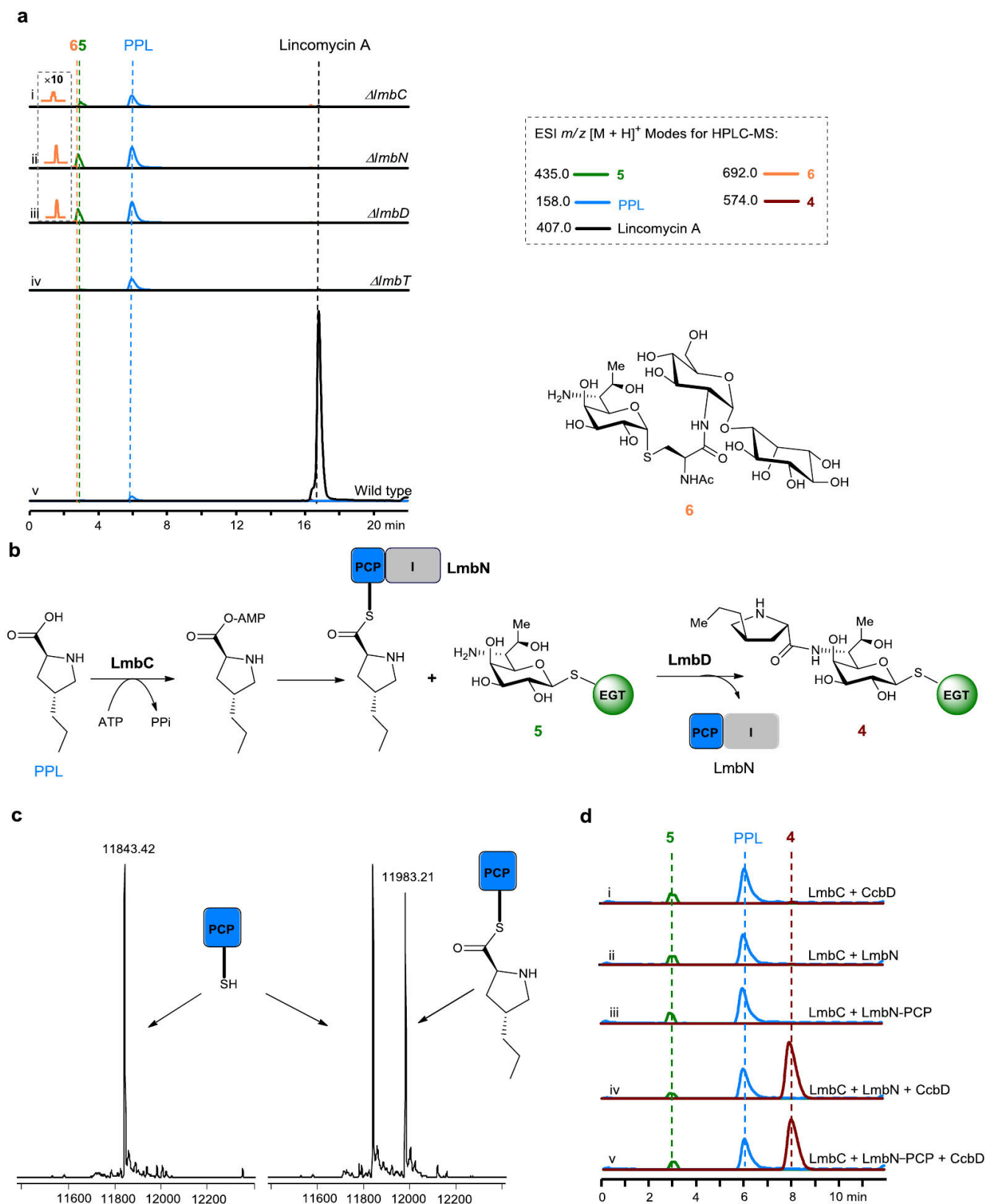
Extended Data Figure 8 | Characterization of PPL incorporation in lincomycin A biosynthesis. For HPLC-MS analysis, the ESI m/z $[M + H]^+$ modes are indicated in the dashed rectangle. **a,** *In vivo* product profiles of *S. lincolnensis* strains, including the mutants (i, for *ΔlmbC*; ii, for *ΔlmbN*; iii, for *ΔlmbD*; and iv, for *ΔlmbT*) and the wild-type control (v). **b,** Process of the incorporation of PPL (with EGT *S*-conjugate **5**) into intermediate **4**. PCP (blue), peptidyl carrier protein; and I (grey), isomerase. **c,** HPLC-MS

examination of LmbC-catalysed conversion of holo-LmbN-PCP (m/z $[M + H]^+$ calculated 11,842.94, found 11,843.42) into PPL-acylated LmbN-PCP (m/z $[M + H]^+$ calculated 11,982.01, found 11,983.21) in the absence (left) and in the presence (right) of ATP. **d,** *In vitro* analysis of condensation between PPL and **5** to generate **4**. The catalyst systems included LmbC + CcbD (i), LmbC + LmbN (ii), LmbC + LmbN-PCP (iii), LmbC + LmbN + CcbD (iv), and LmbC + LmbN-PCP + CcbD (v).

a

b

$K_m = 1.83 \pm 0.18$ mM
$k_{cat} = 6.94 \pm 0.31$ min^{-1}
$k_{cat}/K_m = 3.8$ mM^{-1}min^{-1}

c

$K_m = 0.60 \pm 0.10$ mM
$k_{cat} = 5.19 \pm 0.41$ min^{-1}
$k_{cat}/K_m = 8.7$ mM^{-1}min^{-1}

d
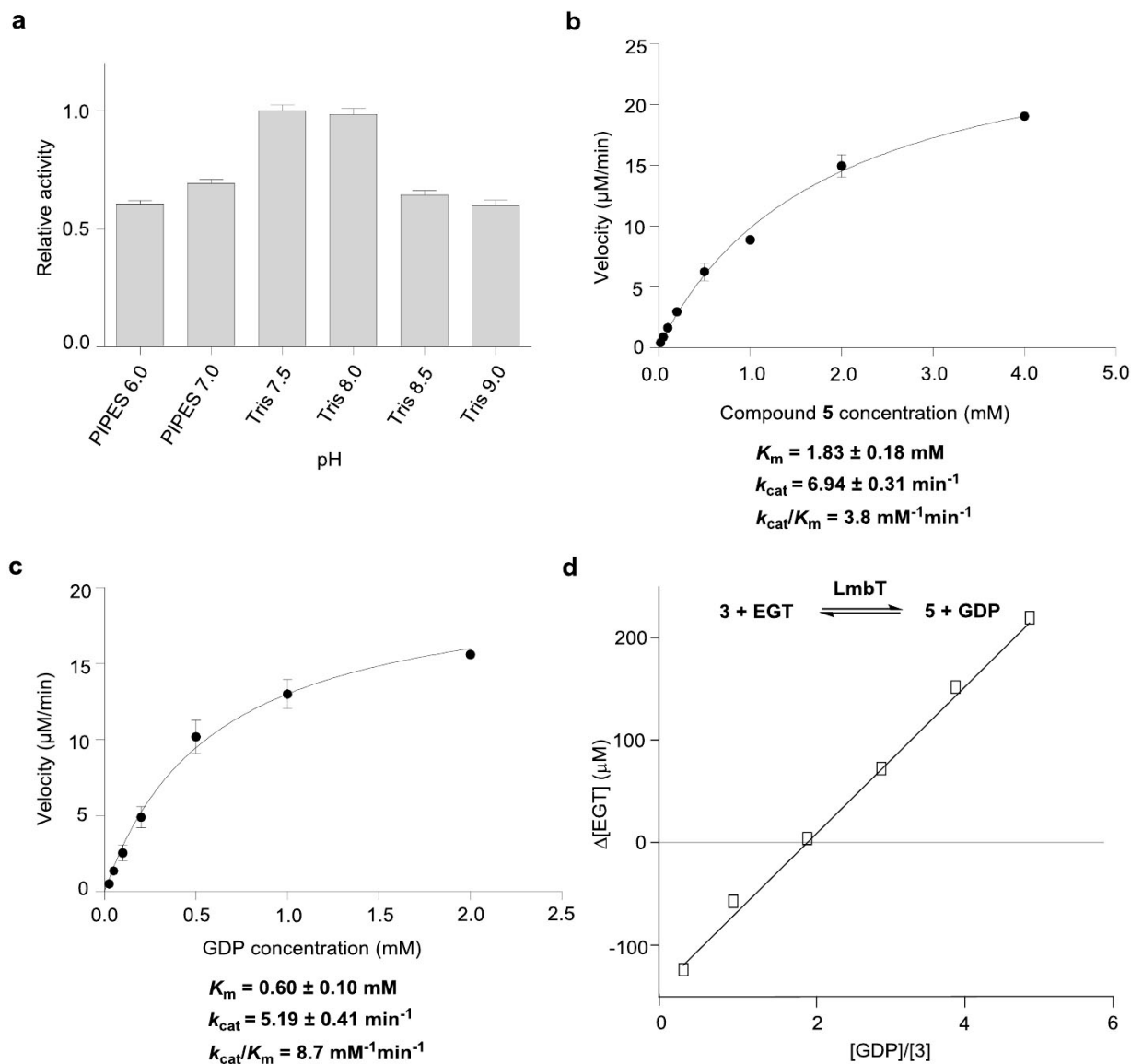
Extended Data Figure 9 | Kinetic analysis of LmbT-catalysed reversible glycosylation. **a**, pH dependence. The activity of LmbT in 50 mM PIPES (pH 6.0–7.0) or 50 mM Tris-HCl (pH 7.5–9.0) buffer was measured. **b**, **c**, Determination of the steady-state kinetic parameters for substrate **5** and for GDP, respectively, in LmbT-catalysed reverse glycosylation. The error bars are standard error of mean ($n = 3$). **d**, Determination of the equilibrium constant (K_{eq}) of LmbT-catalysed glycosylation. $K_{eq} = ([\text{GDP}]/[\mathbf{3}]) \times ([\mathbf{5}]/[\text{EGT}]) = 1.94 \times 1 = 1.94$.
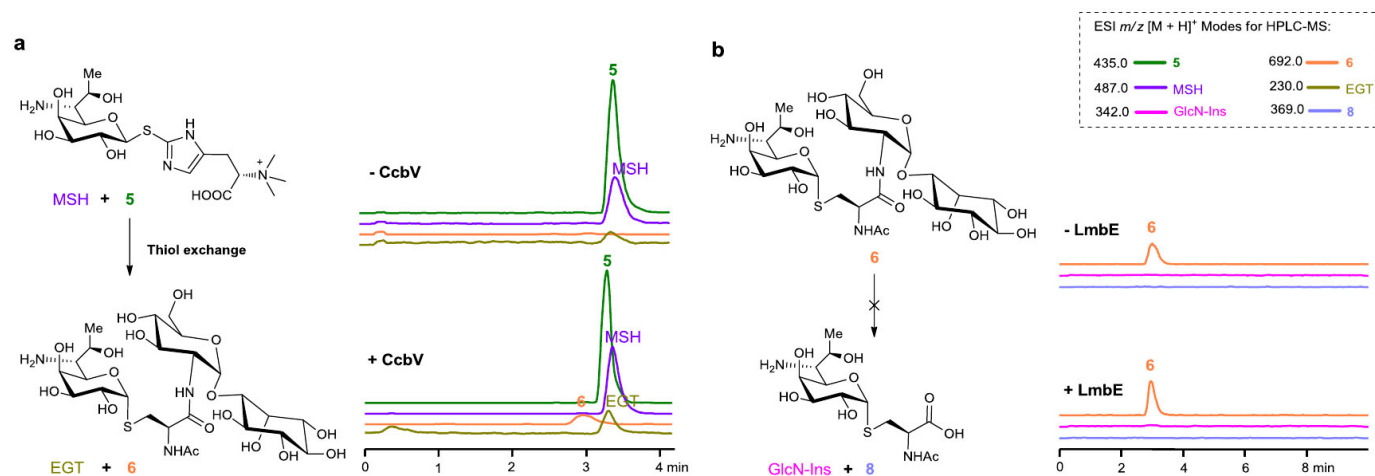
Extended Data Figure 10 | Validation of the highly ordered process involving EGT-mediated assembly and MSH-associated post-modifications in lincomycin biosynthesis. For HPLC-MS analysis, the ESI *m/z* [M + H]$^+$ modes are indicated in the dashed rectangle. **a**, Examination of the thiol exchange using **5** as a substrate in the absence (top right) and in the presence (bottom right) of CcbV, showing that CcbV was able to convert **5** to MSH *S*-conjugate **6** (left). **b**, Determination of the hydrolysis reaction in the absence (top right) and in the presence (bottom right) of LmbE, showing that this enzyme was unable to utilize **6** as a substrate and convert it to the predicted mercapturic acid **8** (left).

LETTER

Structure and function of a single-chain, multi-domain long-chain acyl–CoA carboxylase

Timothy H. Tran[1], Yu-Shan Hsiao[2]†, Jeanyoung Jo[1], Chi-Yuan Chou[1]†, Lars E. P. Dietrich[1], Thomas Walz[2,3] & Liang Tong[1]

Biotin-dependent carboxylases are widely distributed in nature and have important functions in the metabolism of fatty acids, amino acids, carbohydrates, cholesterol and other compounds[1–6]. Defective mutations in several of these enzymes have been linked to serious metabolic diseases in humans, and acetyl-CoA carboxylase is a target for drug discovery in the treatment of diabetes, cancer and other diseases[7–9]. Here we report the identification and biochemical, structural and functional characterizations of a novel single-chain (120 kDa), multi-domain biotin-dependent carboxylase in bacteria. It has preference for long-chain acyl-CoA substrates, although it is also active towards short-chain and medium-chain acyl-CoAs, and we have named it long-chain acyl-CoA carboxylase. The holoenzyme is a homo-hexamer with molecular mass of 720 kDa. The 3.0 Å crystal structure of the long-chain acyl-CoA carboxylase holoenzyme from *Mycobacterium avium* subspecies *paratuberculosis* revealed an architecture that is strikingly different from those of related biotin-dependent carboxylases[10,11]. In addition, the domains of each monomer have no direct contact with each other. They are instead extensively swapped in the holoenzyme, such that one cycle of catalysis involves the participation of four monomers. Functional studies in *Pseudomonas aeruginosa* suggest that the enzyme is involved in the utilization of selected carbon and nitrogen sources.

The reactions catalysed by biotin-dependent carboxylases proceed in two steps and involve at least three different protein components (Extended Data Fig. 1). In the first step, a biotin carboxylase (BC) component catalyses the carboxylation of the biotin cofactor, which is covalently linked to the biotin carboxyl carrier protein (BCCP) component. In the second step, the carboxylated biotin translocates to the carboxyltransferase (CT) active site and transfers the carboxyl group to the substrate. In bacteria, acetyl-CoA carboxylase (ACC) has been well characterized as a multi-subunit enzyme, with a BC subunit, a BCCP subunit and two subunits (α and β) for the CT activity (Extended Data Fig. 1). In contrast, ACC is a large (~250 kDa), single-chain, multi-domain enzyme in most eukaryotes (Extended Data Fig. 1), with domains that are homologous to the bacterial subunits. Other members of this family include propionyl-CoA carboxylase (PCC)[10], 3-methylcrotonyl-CoA carboxylase (MCC)[11], pyruvate carboxylase (PC)[12,13] and urea carboxylase (UC)[14] (Extended Data Fig. 1).

By examining the sequence database, we identified a novel single-chain (~120 kDa), multi-domain biotin-dependent carboxylase in bacteria[1]. The enzyme contains a BC domain at the amino terminus, a BCCP domain near the middle, and a CT domain that is homologous to that of ACC and PCC (Fig. 1a and Extended Data Fig. 1). Homologues of this enzyme are found in a large number of Gram-negative and Gram-positive bacteria, such as *Rhodopseudomonas palustris*, *Mycobacterium avium* subspecies *paratuberculosis*[15] and the human pathogen *Pseudomonas aeruginosa* (Extended Data Fig. 2), with highly conserved sequences (Extended Data Fig. 3). These homologues are in fact mis-annotated as PC[16] or carbamoyl-phosphate synthase (CPS) in the database, as was noted in an earlier report[17], probably because they have roughly the same size as

PC and CPS. The CT domain of PC has a completely different sequence and structure[12,13] (Extended Data Fig. 1), whereas CPS is not a biotin-dependent enzyme and does not have a BCCP domain. These single-chain enzymes are somewhat related to a family of acyl-CoA carboxylases that have been characterized in *Mycobacterium tuberculosis* and other actinomycetes[3], in which BC and BCCP are present in one subunit and CT is in a separate subunit (Extended Data Fig. 1; see below).

We overexpressed several of these single-chain enzymes in *Escherichia coli* and purified them to homogeneity. The proteins migrated at the same position on a gel filtration column as the 750 kDa $\alpha_6\beta_6$ holoenzymes of PCC[10] and MCC[11], suggesting that these enzymes are hexamers, with a molecular mass of ~720 kDa for the holoenzyme.

We characterized the catalytic activities of the enzyme from *R. palustris*, which is annotated as both PC and CPS in the database. Consistent with the sequence analysis was our finding that this enzyme is an acyl-CoA carboxylase, and we did not observe any PC activity for it. The enzyme is active towards all the acyl-CoAs that we tested, with chain lengths from C_2 to C_{16}, but it prefers long-chain substrates (Extended Data Table 1). The k_{cat} values for all the substrates are comparable, while the K_m for palmitoyl-CoA is ~350-fold lower than that for acetyl-CoA. The high sequence conservation among these proteins (Extended Data Fig. 3) suggests that they have similar activity profiles. We have therefore named them long-chain acyl-CoA carboxylases (LCCs). Broad-spectrum activity has been observed for a few of the acyl-CoA carboxylases in actinomycetes[3,18].

To define the holoenzyme architecture, we determined the crystal structure of *M. avium* subspecies *paratuberculosis* LCC (MapLCC) at 3.0 Å resolution (Fig. 1b–e); MapLCC shares 52% amino acid sequence identity with *R. palustris* LCC (Extended Data Fig. 3). The holoenzyme hexamer is situated on a crystallographic three-fold axis, and there is a dimer in the asymmetric unit. The atomic model has good agreement with the crystallographic data and the expected geometric parameters (Extended Data Table 2). Several segments of the protein have poor or no electron density and are not included in the atomic model. These include parts of the linkers from the BCCP to the BC and CT domains (Fig. 1b), although there is no ambiguity in assigning the BCCP domain to a specific monomer. A different assignment will result in gaps that are too large to be bridged by the missing residues.

The overall structures of the two MapLCC monomers in the asymmetric unit are similar. With their CT domains superposed, a difference of 3° is seen in the orientation of their BC domains (Fig. 1b). The BCCP domains have a larger difference, corresponding to a rotation of 15°, indicating some asymmetry in the holoenzyme hexamer. The BC, BCCP and CT domains of each monomer do not have any direct contact with each other.

The structure of the holoenzyme hexamer of MapLCC has the shape of an equilateral triangle, obeying 32 symmetry and with a length of ~180 Å for each side (Fig. 1c) and a thickness of ~65 Å (Fig. 1d, e). A hexamer of the CT domain forms the central core of the structure, with three CT domains in each layer and CT dimers being formed by one domain from each layer. Dimeric BC domains are located at the vertices

---

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present addresses: Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA (Y.-S.H.); Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 112, Taiwan (C.-Y.C.).
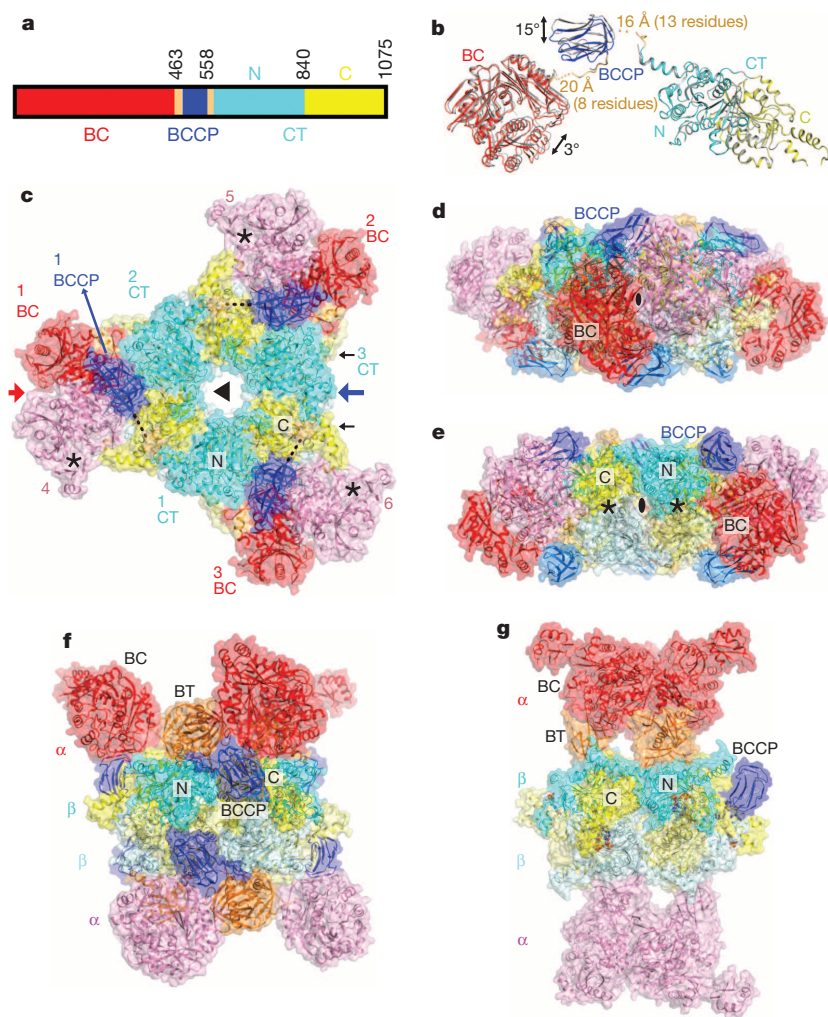
Figure 1 | Crystal structure of LCC from *M. avium* subspecies *paratuberculosis* (MapLCC). a, Domain organization of MapLCC. The domains are labelled and given different colours. **b,** Overlay of the structures of the two MapLCC monomers in the asymmetric unit (one in colour, the other in grey). Residues that are missing in the linkers from BCCP are indicated with dashed lines. **c,** Overall structure of the 720 kDa hexameric holoenzyme of MapLCC. The six monomers are labelled. The domains in the three monomers in the top layer (numbered 1, 2 and 3) are coloured as in **a**. The BC, BCCP, N and C CT domains in the three monomers in the bottom layer (numbered 4, 5 and 6) are coloured pink, pale blue, pale cyan and pale yellow, respectively. The disordered region of the BCCP–CT linker is indicated with a dashed line (black). The BC active sites are indicated with asterisks (black). The CT active sites are on the side of the CT domain core, indicated with black arrows. **d,** Structure of the MapLCC holoenzyme viewed down the BC domain dimer (red arrow in **c**). **e,** Structure of the MapLCC holoenzyme viewed down the blue arrow in **c**. The CT active sites are indicated with asterisks (black). **f,** Structure of the 750 kDa $\alpha_6\beta_6$ PCC holoenzyme[10]. The view is equivalent to that in **d**. **g,** Structure of the 750 kDa $\alpha_6\beta_6$ MCC holoenzyme[11]. The structure figures were produced with PyMOL (http://www.pymol.org).

of the triangle, contacting CT domain dimers. The BCCP domains are situated between the BC and CT domains but not in the active site of either domain. The lysine residue that would be biotinylated is located on the surface, more than 15 Å away from the nearest BC or CT domain, and its side chain is disordered. The BCCP domain is not biotinylated in this holoenzyme, even though it was expressed under conditions identical to those for *R. palustris* LCC, which was completely biotinylated.

The overall architecture of the MapLCC holoenzyme is strikingly different from those of PCC (Fig. 1f)[10] and MCC (Fig. 1g)[11], even though all three enzymes are ~750 kDa oligomers made up of homologous BC, CT and BCCP domains. The central CT domain core of MapLCC is similar to that of the β_6 hexamer core of PCC, with a root mean squared distance of 1.7 Å for 2,073 equivalent Cα atoms between them (Extended Data Fig. 4). In contrast, whereas the BC domains are located above and below the central core in PCC and MCC, they are positioned at the side of the central core in MapLCC. Moreover, the BC domain is a monomer in PCC and a weakly associated trimer in MCC, but it is a dimer in MapLCC. In fact, this dimer is similar to that for the BC subunit of *E. coli* ACC[19,20] (Extended Data Fig. 4) and the BC domain of PC[12,13].

A BT domain was identified in the structures of PCC (Fig. 1f)[10] and MCC (Fig. 1g)[11], located between the BC and BCCP domains in the primary sequence (Extended Data Fig. 1) and having an important function in mediating interactions between their α (BC) and β (CT) subunits[10,11]. This domain does not exist in LCC, because there are insufficient residues in the linker between BC and BCCP to form such a domain (Extended Data Fig. 1). However, the BC–BCCP linker and the BCCP–CT linker do participate in mediating interactions in the MapLCC holoenzyme, and there are also direct contacts between the BC and CT domains (see below).

The domains of the MapLCC monomers are swapped extensively in the holoenzyme hexamer. The CT domains of two monomers related by a BC domain dimer are located far from each other (Fig. 2a). Similarly, the BC domains of two monomers related by a CT domain dimer are also located far from each other (Fig. 2b). Interactions between these domains therefore occur only in the context of the holoenzyme, and the structure of the monomer alone is unlikely to be stable (Fig. 1b).

A total of ~8,200 Å² of the surface area of each MapLCC monomer is buried in the holoenzyme. The majority of this surface, 5,700 Å², is buried by the interface among the CT domains in the central core, and a 1,000 Å² surface is buried in the BC dimer interface (Fig. 3a, b). The remaining interfaces in the holoenzyme make smaller contributions to the surface area burial, 600 Å² in the BC–CT interface, 500 Å² for the linkers from BCCP, and 400 Å² from the BCCP domain. Nonetheless, all of these smaller interfaces together may be important for stabilizing the holoenzyme.

The primary BC–CT interface involves three monomers of MapLCC. For example, a β-hairpin structure in the BC domain of monomer 4 contacts the CT domains of monomer 2 (C domain) and monomer 6 (N domain; Fig. 3c). Similarly, the BC–BCCP linker just before the BCCP domain in monomer 1 contacts the CT domain of monomer 2 and the BC domain of monomer 4 (Fig. 3d). These interactions also demonstrate the extensive connections between the monomers in the holoenzyme. The interfaces involve van der Waals contacts, hydrogen bonding and ionic interactions.

We also carried out electron microscopy (EM) studies on the MapLCC holoenzyme. Negative-stain EM images showed monodisperse particles of similar sizes but variable shapes (Extended Data Fig. 5). Classification
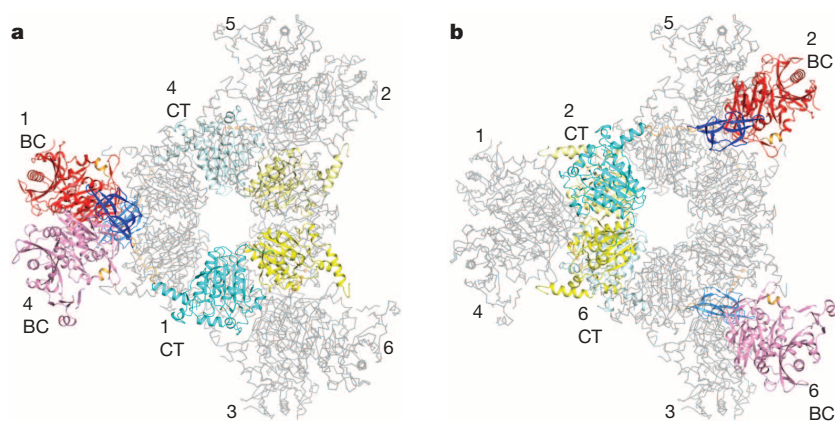
Figure 2 | Extensive domain swapping in the structure of the MapLCC holoenzyme. a, The CT domains of two monomers (in colour) related by a BC dimer are located far from each other (note that they are in opposite layers) and participate in dimerization with different monomers (in grey). The six monomers are labelled. **b**, The BC domains of two monomers related by a CT dimer are located far from each other.

in SPARX[21] of ~25,000 particles yielded 308 classes that represented 65% of the data set (Extended Data Fig. 5). Cross-correlation analysis identified class averages very similar to the top view (cross-correlation coefficient 0.856; Fig. 3e) and side view (cross-correlation coefficient 0.752; Fig. 3f) of the crystal structure, confirming the holoenzyme architecture seen in the crystal. At the same time, substantial variations in the three peripheral densities (corresponding to the BC domains) relative to the central core (the CT domains) are also observed (Extended Data Fig. 5). Each peripheral density can appear as a single or bilobed feature (Fig. 3g), probably representing different views of the BC dimer. The densities can also be located in different positions relative to the central core, breaking the alignment of the two-fold axes of the BC and CT dimers (Fig. 3h). Overall, the EM data indicate that the peripheral BC domains move as dimers and are flexibly tethered to the CT core (Supplementary Video 1), consistent with the crystal structure that shows that the BC–CT contact is relatively weak in MapLCC.

Each BC active site is located ~40 Å from a CT active site in the MapLCC holoenzyme (Fig. 4a). The BCCP domain, although not located in either active site, can access both of them, through conformational changes in its linkers. The crystal structure shows that the BCCP domain of monomer 1 visits the BC active site of monomer 4 and the CT active site at the interface of monomers 2 and 6 (Fig. 4a), suggesting that each cycle of catalysis requires the participation of four monomers. This again indicates the extensive communications between the monomers in this holoenzyme. The residues in both active sites are generally conserved with those in other biotin-dependent carboxylases, suggesting a similar catalytic mechanism[22] for LCC. The CT active site has a pocket that can accommodate short-chain and medium-chain acyl groups (Extended Data Fig. 6), and a conformational change is needed to bind long-chain substrates. This conformational flexibility may be important for the enzyme to adapt to and be active towards the broad collection of substrates.

In most of the other biotin-dependent carboxylases, BCCP is located at the end of a polypeptide chain, and there is therefore only one linker to the rest of the protein. The LCCs, and the eukaryotic ACCs (Extended Data Fig. 1), are distinct in that BCCP is located in the middle of those proteins. There are therefore two linkers from BCCP to the rest of the protein. The amino acid sequences of these two linkers in the LCC enzymes are not conserved (Extended Data Fig. 3). They are therefore likely to be flexible and allow the movement of BCCP during catalysis, which is consistent with the fact that a portion of both linkers is disordered in the structure (Fig. 1b). Residues in both linkers that are included in the current atomic model also have high *B* values (Extended Data Fig. 4). In addition, the B domain of BC and a loop near the carboxy terminus of the protein have high *B* values and are partly disordered.

The structure of MapLCC also has implications for the holoenzymes of other biotin-dependent carboxylases, especially the family of acyl-CoA carboxylases in *M. tuberculosis, S. coelicolor* and other actinomycetes[3]. These enzymes contain two subunits, with BC–BCCP in the α subunit

and CT in the β subunit, and there are not enough residues in one of the α subunits for a BT domain (Extended Data Fig. 1). The holoenzyme is an $\alpha_6\beta_6$ dodecamer, and the structure of the β_6 hexamer[23] is similar to that of the CT domain hexamer in MapLCC. It is therefore likely that the holoenzymes of such two-subunit carboxylases (lacking the BT domain) share a similar architecture to that of MapLCC (Fig. 1c) rather than PCC (Fig. 1f). *M. tuberculosis* does not have an LCC homologue, indicating some differences between these mycobacterial species.

Eukaryotic ACCs are also single-chain, multi-domain enzymes (Extended Data Fig. 1), although there are substantial differences from LCC. The eukaryotic ACCs contain ~1,000 additional residues, including a unique central region of ~700 residues, and they are likely to carry a BT domain as well. The overall architecture of the eukaryotic ACCs is therefore likely to be different from that of MapLCC observed here.

We have begun to characterize the physiological functions of LCC, using *P. aeruginosa* (strain PA14) as the model organism. The *P. aeruginosa* enzyme (locus name PA14_46320, homologue of PA1400 in *P. aeruginosa* PAO1) shares 59% amino acid sequence identity with *R. palustris* LCC. This organism also carries the multi-subunit ACC, MCC and geranyl-CoA carboxylase (GCC)[24,25], although it lacks a homologue for PCC. The multi-subunit ACC is probably essential, in a similar manner to the *E. coli* enzyme, because no transposon insertions in its subunits are found in a PA14 transposon mutant library[26]. In contrast, transposon insertions are found in the loci encoding LCC, MCC and GCC, suggesting that they are not essential for growth under the conditions used to produce this library[26].

We produced a markerless deletion of LCC/PA14_46320, confirming that it is not essential for cell survival. The activity profiles of the mutant under ~2,000 conditions were characterized by using a phenotype microarray, which monitored its ability to reduce a tetrazolium dye[27]. The incubation conditions included different carbon or nitrogen sources, nutrient supplements, osmolytes, pH values and antibiotics. Although most of the conditions showed comparable profiles between the wild type and the LCC deletion mutant, phenotypic differences were observed for several conditions (Extended Data Fig. 7), and two of these were considered significant on the basis of the phenotype microarray analysis—using fumarate as the sole carbon source and the Met-Val dipeptide as the sole nitrogen source (Fig. 4b). It is not clear how this LCC enzyme is linked to the utilization of selected carbon and nitrogen sources in *P. aeruginosa*. The metabolism of both Met and Val is likely to produce propionyl-CoA, and the PCC activity of LCC may be important for its further degradation, as is the case in many other organisms. This is also supported by the fact that *P. aeruginosa* lacks a proper PCC enzyme.

A long-chain acyl-CoA carboxylase activity is needed for the biosynthesis of mycolic acid in *Mycobacterium* and other actinomycetes[3,18,28]. However, the LCC enzyme studied here is unlikely to be essential for this function, because its homologue is absent in *M. tuberculosis*. Moreover, mycolic acid is not known to be present in *P. aeruginosa*, *R. palustris* and other bacteria, suggesting that this LCC is likely to have different physiological functions.
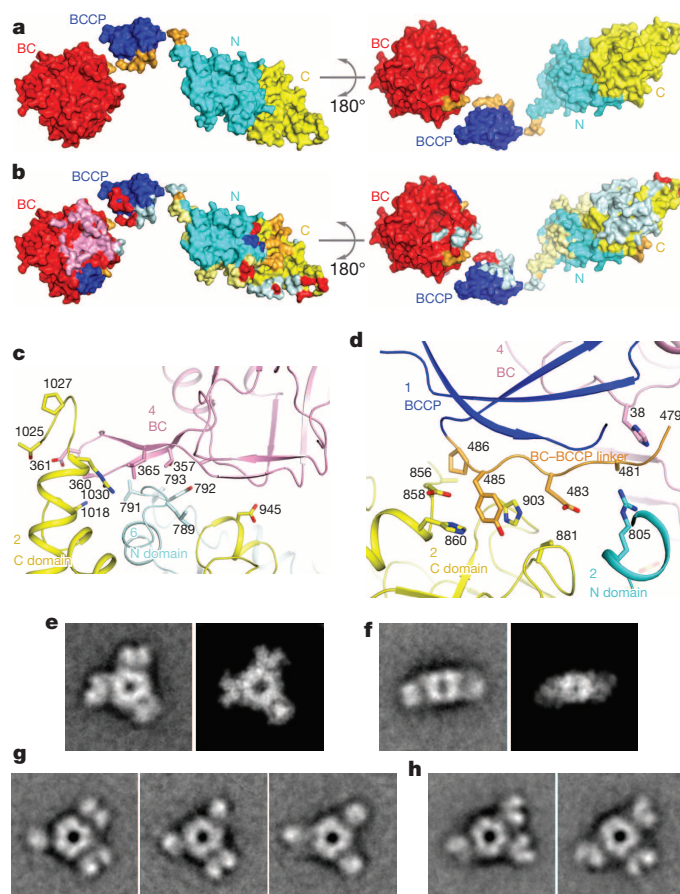
Figure 3 | Interactions between the central core and the rest of the holoenzyme. a, Two views of the molecular surface of a MapLCC monomer, coloured by the domains. **b**, Residues that mediate interactions in the MapLCC holoenzyme are indicated by the colours of the domains that they contact. For example, a pale cyan patch on the surface of the C domain of CT indicates residues that interact with the N domain of another monomer in the holoenzyme. The pink patch on BC indicates residues in the BC dimer interface, and the pale cyan patch indicates residues in the interface with CT. The major areas of contacts are between the CT domains and in the BC domain dimer. The views are identical to those in **a**. **c**, Residues in the interface between the BC and CT domains. The domains and the monomers they belong to are labelled. **d**, Residues in the interface between the BC–BCCP linker just before the BCCP domain and the rest of the holoenzyme. **e**, Class average representing the top view (left) most similar to the crystal structure and the corresponding projection from the crystal structure filtered to 30 Å resolution (right). **f**, Class average representing the side view (left) most similar to the crystal structure and the corresponding projection from the crystal structure filtered to 30 Å resolution (right). **g**, Three panels showing that each of the peripheral BC domains can appear as a single or bilobal density, probably representing different orientations of the BC dimer. **h**, Two panels illustrating that the BC domains can adopt different positions around the central CT core. The side length of the individual panels in **e**–**h** is 340 Å.

The striking differences between the architectures of the LCC, PCC and MCC holoenzymes are also linked to functional differences between them. Especially, the N and C domains of the β subunits of PCC and MCC are swapped relative to each other, and this is coupled to the different substrate specificity of the two enzymes. PCC carboxylates the α carbon of an acid (as a CoA ester), whereas MCC carboxylates the γ carbon of an α-β unsaturated acid (Extended Data Fig. 1). The structural differences therefore suggest that there are two lineages of the biotin-dependent carboxylases[11], one containing PCC, ACC and LCC, and the other containing MCC and possibly GCC. At the same time, PCC and LCC share propionyl-CoA carboxylase activity, indicating that similar biochemical activities can be supported by holoenzymes with markedly different architectures as well.
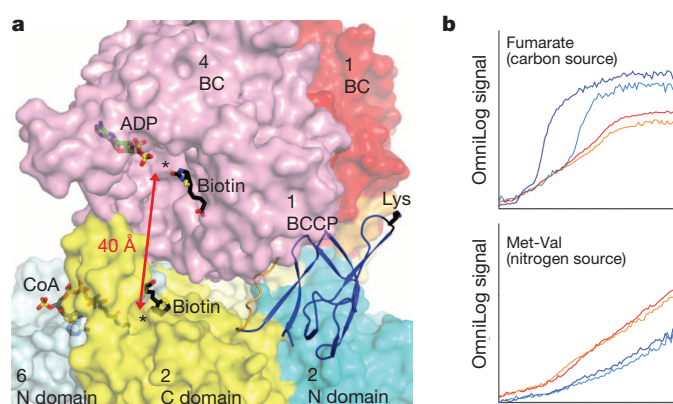
Figure 4 | Catalysis and function of LCC. a, The BC and CT active sites (black asterisks) of MapLCC are separated by ~40 Å (red arrow). Molecular surface of MapLCC is shown, coloured as in Fig. 1c. The domains are labelled by the monomer they belong to. The Lys side chain of BCCP to which biotin would be connected to is shown in black. The binding modes of ADP (green) and biotin (black) to the *E. coli* BC subunit are shown as stick models[20]. The binding mode of CoA (grey) to the CT domain of yeast ACC[29] and biotin (black) in the β (CT) subunit of PCC[10] are also shown. **b**, Phenotypic differences between wild-type and LCC knockout (ΔPA14_46320) *P. aeruginosa* strains, revealed by a colorimetric assay that monitors the reduction of a tetrazolium dye. Assays were performed twice in each medium for the wild-type (red and orange) and mutant (blue and cyan) strains. For each panel, the horizontal axis is time (24 h) and the vertical axis is OmniLog signal[27].

Overall, our studies have identified a new member of the biotin-dependent carboxylase family, and revealed a novel architecture for its holoenzyme. These observations are also relevant for other members of this family. Moreover, the differences between the architectures of LCC, PCC and MCC holoenzymes, despite their sharing domains with homologous structures, indicate that these domains can be arranged in remarkably different ways to form the various holoenzymes. This may have substantial implications for other multi-domain proteins, especially eukaryotic ACCs, and for protein structure and sequence conservation in general.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Tong, L. Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* **70,** 863–891 (2013).
2. Waldrop, G. L., Holden, H. M. & St. Maurice, M. The enzymes of biotin dependent CO_2 metabolism: what structures reveal about their reaction mechanisms. *Protein Sci.* **21,** 1597–1619 (2012).
3. Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.-C. & Gramajo, H. Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol. Rev.* **35,** 475–497 (2011).
4. Jitrapakdee, S. *et al.* Structure, mechanism and regulation of pyruvate carboxylase. *Biochem. J.* **413,** 369–387 (2008).
5. Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell. Mol. Life Sci.* **62,** 1784–1803 (2005).
6. Cronan, J. E. Jr & Waldrop, G. L. Multi-subunit acetyl-CoA carboxylases. *Prog. Lipid Res.* **41,** 407–435 (2002).
7. Polyak, S. W., Abell, A. D., Wilce, M. C. J., Zhang, L. & Booker, G. W. Structure, function and selective inhibition of bacterial acetyl-CoA carboxylase. *Appl. Microbiol. Biotechnol.* **93,** 983–992 (2012).
8. Abramson, H. N. The lipogenesis pathway as a cancer target. *J. Med. Chem.* **54,** 5615–5638 (2011).
9. Wakil, S. J. & Abu-Elheiga, L. A. Fatty acid metabolism: target for metabolic syndrome. *J. Lipid Res.* **50,** S138–S143 (2009).
10. Huang, C. S. *et al.* Crystal structure of the $\alpha_6\beta_6$ holoenzyme of propionyl-coenzyme A carboxylase. *Nature* **466,** 1001–1005 (2010).
11. Huang, C. S., Ge, P., Zhou, Z. H. & Tong, L. An unanticipated architecture of the 750-kDa $\alpha_6\beta_6$ holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature* **481,** 219–223 (2012).
12. St. Maurice, M. *et al.* Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* **317,** 1076–1079 (2007).
13. Xiang, S. & Tong, L. Crystal structures of human and *Staphylococcus aureus* pyruvate carboxylase and molecular insights into the carboxyltransferase reaction. *Nature Struct. Mol. Biol.* **15,** 295–302 (2008).

14. Fan, C., Chou, C.-Y., Tong, L. & Xiang, S. Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J. Biol. Chem.* **287,** 9389–9398 (2012).
15. Li, L. *et al.* The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis. Proc. Natl Acad. Sci. USA* **102,** 12344–12349 (2005).
16. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406,** 959–964 (2000).
17. Lai, H., Kraszewski, J. L., Purwantini, E. & Mukhopadhyay, B. Identification of pyruvate carboxylase genes in *Pseudomonas aeruginosa* PAO1 and development of a *P.aeruginosa*-based overexpression system for a4- and α₄β₄-type pyruvate carboxylases. *Appl. Environ. Microbiol.* **72,** 7785–7792 (2006).
18. Grande, R. *et al.* The two carboxylases of *Corynebacterium glutamicum* essential for fatty acid and mycolic acid synthesis. *J. Bacteriol.* **189,** 5257–5264 (2007).
19. Waldrop, G. L., Rayment, I. & Holden, H. M. Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33,** 10249–10256 (1994).
20. Chou, C.-Y., Yu, L. P. C. & Tong, L. Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism. *J. Biol. Chem.* **284,** 11690–11697 (2009).
21. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
22. Knowles, J. R. The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* **58,** 195–221 (1989).
23. Lin, T. W. *et al.* Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium tuberculosis. Proc. Natl Acad. Sci. USA* **103,** 3072–3077 (2006).
24. Forster-Fromme, K. & Jendrossek, D. Catabolism of citronellol and related acyclic terpenoids in pseudomonads. *Appl. Microbiol. Biotechnol.* **87,** 859–869 (2010).
25. Aguilar, J. A. *et al.* Substrate specificity of the 3-methylcrotonyl coenzyme A (CoA) and geranyl-CoA carboxylases from *Pseudomonas aeruginosa. J. Bacteriol.* **190,** 4888–4893 (2008).
26. Liberati, N. T. *et al.* An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci. USA* **103,** 2833–2838 (2006).
27. Shea, A., Wolcott, M., Daefler, S. & Rozak, D. A. Biolog phenotype microarrays. *Methods Mol. Biol.* **881,** 331–373 (2012).
28. Takayama, K., Wang, C. & Besra, G. S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis. Clin. Microbiol. Rev.* **18,** 81–101 (2005).
29. Zhang, H., Yang, Z., Shen, Y. & Tong, L. Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299,** 2064–2067 (2003).

Author Contributions T.H.T. and C.-Y.C. performed cloning, protein expression, purification and crystallization experiments. T.H.T. and L.T. performed the crystallography experiments and calculation. Y.-S.H. and T.W. conducted the electron microscopy experiments. T.H.T. and C.-Y.C. performed the kinetic assays. J.J. and L.E.P.D. conducted the *P. aeruginosa* experiments. T.H.T., L.E.P.D., T.W. and L.T. wrote the paper.

Author Information Coordinates and structure factors have been deposited in Protein Data Bank under accession number 4RCN. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).

METHODS

Protein expression and purification. Full-length LCCs from several different bacterial organisms, including *R. palustris*, *M. avium* subspecies *paratuberculosis* and *P. aeruginosa*, were amplified from genomic DNA by PCR and cloned into pET28a, pET26b and/or pET24d vectors (Novagen). The plasmids were transformed into BL21Star (DE3) cells (Invitrogen). Protein expression was induced with the addition of 1 mM isopropyl β-D-thiogalactoside, and the cells were grown at 16 °C for 16–20 h.

To facilitate biotinylation, the recombinant enzyme was co-expressed with the *E. coli* biotin ligase BirA, and 15 mg l^{-1} biotin was added to the medium. An avidin shift assay of the purified enzymes showed that *R. palustris* LCC was completely biotinylated. However, purified MapLCC did not show any biotinylation, possibly indicating some degree of selectivity of the BirA enzyme. Expression of *P. aeruginosa* LCC did not produce any soluble protein and was not pursued further.

Cells were lysed by sonication in a buffer containing 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% (v/v) glycerol, 10 mM 2-mercaptoethanol and 1 mM phenyl-methylsulphonyl fluoride. Soluble enzyme was purified by Ni^{2+}-nitrilotriacetate (Qiagen), anion-exchange and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. The S-300 running buffer for MapLCC contained 25 mM HEPES pH 7.4, 250 mM NaCl and 2.5 mM dithiothreitol. The purified protein was concentrated to 6 mg ml^{-1}, and the solution was supplemented with 5% (v/v) glycerol before being flash-frozen in liquid nitrogen and stored at −80 °C.

The selenomethionyl MapLCC protein was produced in B834 (DE3) cells (Novagen) that were grown in defined LeMaster medium supplemented with selenomethionine[30]. The protein was purified with the same protocol as that for the native enzyme.

Protein crystallization. MapLCC was crystallized at 4 °C using the microbatch method under paraffin oil. The protein solution was mixed with a precipitant solution containing 0.1 M Bis-Tris propane pH 7.5–8.5 and 1.5–2.0 M ammonium sulphate. Crystals took 4–6 weeks to grow to full size, and larger crystals were obtained by microseeding. They were cryoprotected with reservoir solution supplemented with 12–15% (v/v) glycerol and flash-frozen in liquid nitrogen for data collection at 100 K. The C-terminal His tag on the protein was not removed before crystallization.

Data collection and structure determination. X-ray diffraction data for the native (wavelength 1.075 Å) and selenomethionyl (0.979 Å) crystals were collected with a Q315 charge-coupled device (Area Detector Systems Corporation) at the X29A beamline of the National Synchrotron Light Source. The diffraction images were processed with the HKL package[31]. The crystals belong to space group $P2_13$, with cell dimensions of $a = b = c = 220.9$ Å. There are two MapLCC monomers in the crystallographic asymmetric unit.

The structure of MapLCC was solved by a combination of molecular replacement and selenomethionyl SAD phasing. The orientation and position of the BC, CT and BCCP domains were located with the program Phaser[32]. The Se sites were located with the program SHELX[33], and SOLVE/RESOLVE was used for phasing the reflections and automated model building[34]. The atomic model was built with the program Coot[35]. The structure refinement was performed with the program CNS[36]. The crystallographic information is summarized in Extended Data Table 2.

We also obtained a second crystal form of MapLCC, with an entire hexamer in the asymmetric unit, and were able to collect an X-ray diffraction data set to 4.3 Å resolution (space group $P2_12_12_1$, $a = 102$ Å, $b = 292$ Å, $c = 314$ Å). The structure of this crystal form was readily solved by the molecular replacement method, and it revealed essentially the same holoenzyme architecture (data not shown).

Electron microscopy and image processing. Purified MapLCC was prepared by conventional negative staining with 0.75% (w/v) uranyl formate[37]. Images were collected with a Tecnai T12 electron microscope (FEI) equipped with an LaB$_6$ filament and operated at an acceleration voltage of 120 kV. Images were recorded using low-dose procedures on an UltraScan 895 4K x 4K charge-coupled device (CCD) camera (Gatan) using a defocus of −1.5 μm and a nominal magnification of ×52,000. The calibrated magnification was ×70,527, yielding a pixel size of 2.13 Å on the specimen level.

BOXER, the display program associated with the EMAN software package[38], was used to select 24,535 particles interactively from 270 CCD images, and the SPIDER software package[39] was used to window the particles into 160 × 160-pixel images. To perform iterative stable alignment and clustering (ISAC)[40] in SPARX[21], the size of the particle images was reduced to 64 × 64 pixels, and the particles were pre-aligned and centred. ISAC was run on the Orchestra High Performance Compute Cluster at Harvard Medical School (http://rc.hms.harvard.edu), specifying 200 images per group and a pixel error of 0.7. After 19 generations, 308 classes were obtained, accounting for 15,932 particles (65% of the entire data set) (Extended Data Fig. 5). Averages of these classes were calculated using the original 160 × 160-pixel images.

To confirm that the ISAC averages were representative of the entire data set, the particles were also subjected to ten cycles of multi-reference alignment in SPIDER. Each round of multi-reference alignment was followed by K-means classification, specifying 300 output classes (Extended Data Fig. 5).

To compare the class averages with the crystal structure, the crystal structure was Fourier transformed, filtered to 30 Å with a Butterworth low-pass filter, and transformed back. Evenly spaced projections were calculated at 4° intervals and subjected to ten cycles of alignment with masked EM class averages. The class averages most similar to the top and side views of the crystal structure and their cross-correlation coefficients are presented in Fig. 3e, f.

To visualize the structural heterogeneity of MapLCC in solution, 179 averages obtained by ISAC that showed three peripheral densities were selected (indicated by asterisks in Extended Data Fig. 5), ordered according to their correlation coefficients, and used to prepare Supplementary Video 1. This structural variability is probably the reason why it was not possible to calculate a three-dimensional map from cryo-EM images of vitrified MapLCC samples.

Enzymatic assays. The kinetic assays monitored the hydrolysis of ATP by *R. palustris* LCC in the presence of various acyl-CoA substrates, using coupling enzymes to convert the ADP product to NADH oxidation[41]. The reaction mixture contained 100 mM HEPES pH 7.5, 40 mM KHCO$_3$, 1.5 mM ATP, 0.4 mM NADH, 200 mM KCl, 10 mM MgCl$_2$, 0.5 mM phosphoenolpyruvate, 3.5/3.7 U of lactate dehydrogenase/pyruvate kinase (Sigma), 0.25 μM enzyme (except for MCC, which was at 1.2 μM) and various concentrations of acyl-CoA. The absorbance at 340 nm was monitored for 1.5 min. The initial velocities were fitted to the Michaelis–Menten equation using the program Origin (OriginLab).

Construction of an LCC deletion mutant in *P. aeruginosa*. A markerless deletion was generated for the gene PA14_46320 in *P. aeruginosa* PA14, using previously described methods[42]. In brief, ~1-kilobase flanking regions for PA14_46320 were amplified with primers listed in Extended Data Table 3 and recombined into the allelic-replacement vector pMQ30 through gap repair cloning in the yeast strain InvSc1 (ref. 43). This plasmid was transformed into *E. coli* BW29427 and moved into PA14 using biparental conjugation. Luria–Bertani (LB) agar containing 100 μg ml^{-1} gentamicin was used to select for *P. aeruginosa* single recombinants. Markerless deletions in PA14_46320 (double recombinants) were then selected with the use of LB agar plates devoid of NaCl and containing 10% (w/v) sucrose as a counter-selection, and their genotypes were confirmed by PCR.

Phenotype microarrays. Phenotype microarray screening was performed by Biolog, Inc., as described[27].

30. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).
31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
33. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
34. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).
35. Emsley, P. & Cowtan, K. D. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
36. Brunger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol. Proced. Online* **6**, 23–34 (2004).
38. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
39. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
40. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
41. Blanchard, C. Z., Lee, Y. M., Frantom, P. A. & Waldrop, G. L. Mutations at four active site residues of biotin carboxylase abolish substrate-induced synergism by biotin. *Biochemistry* **38**, 3393–3400 (1999).
42. Recinos, D. A. *et al.* Redundant phenazine operons in *Pseudomonas aeruginosa* exhibit environment-dependent expression and differential roles in pathogenicity. *Proc. Natl Acad. Sci. USA* **109**, 19420–19425 (2012).
43. Shanks, R. M., Caiazza, N. C., Hinsa, S. M., Toutain, C. M. & O'Toole, G. A. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from Gram-negative bacteria. *Appl. Environ. Microbiol.* **72**, 5027–5036 (2006).
44. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
45. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
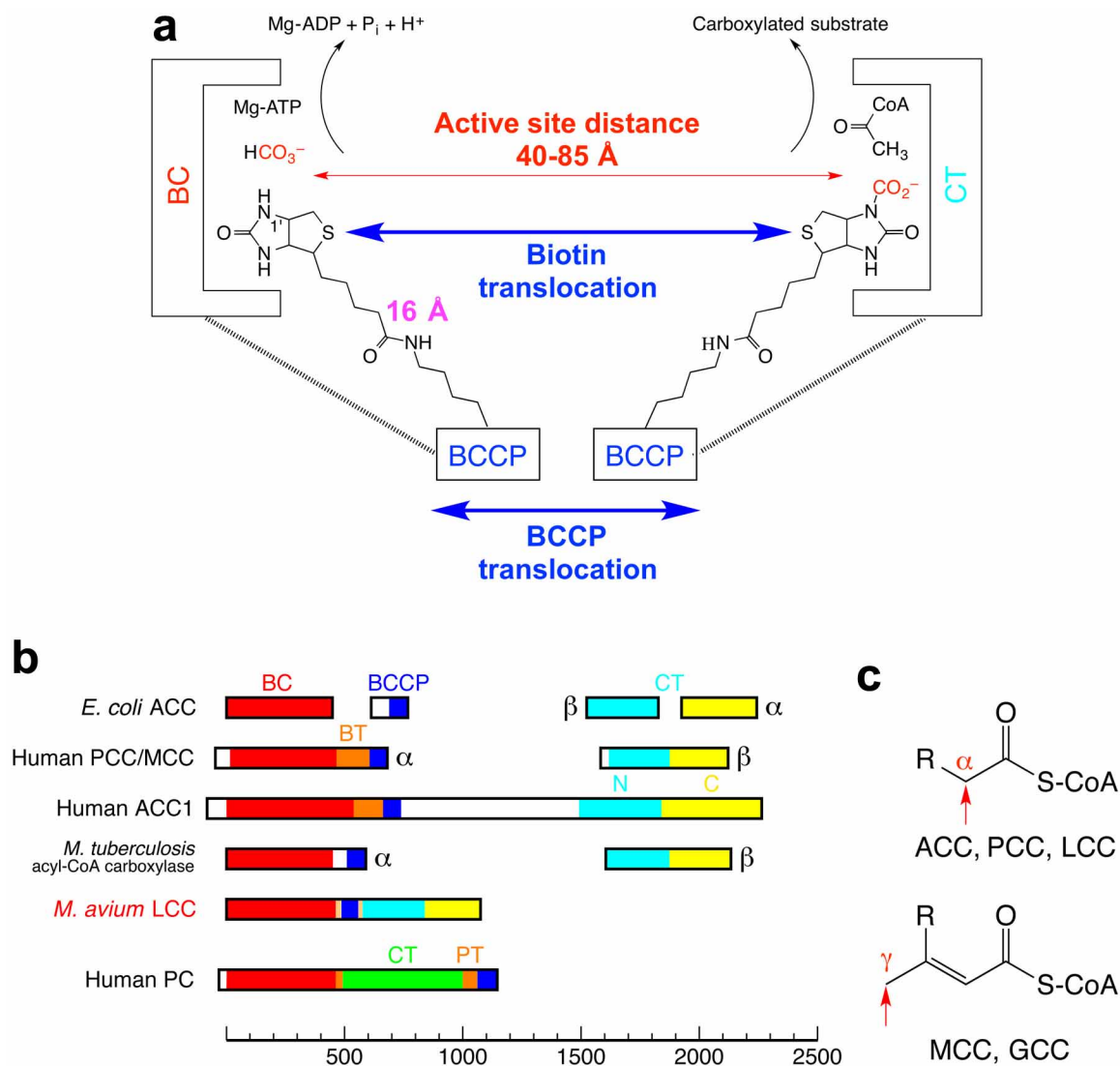
Extended Data Figure 1 | Domain organization of biotin-dependent carboxylases. a, Reactions catalysed by biotin-dependent carboxylases. Biotin is linked to the side chain of a Lys residue in BCCP, and this flexible arm has a maximum length of ~16 Å. The BCCP domain must also translocate to reach both active sites, separated by distances of 40–85 Å based on known holoenzyme structures (swinging domain model). **b**, Domain organizations of several representative biotin-dependent carboxylases. Homologous domains are given the same colours. The CT domain of PC has a completely different sequence and structure from those of ACC and PCC. The proteins are drawn to scale, and a scale bar is shown at the bottom. BT, BC–CT interaction domain; PT, PC tetramerization domain, also known as allosteric domain. **c**, Chemical structures of the substrates of ACC, PCC, LCC, MCC and GCC. The site of carboxylation is indicated with the red arrow.
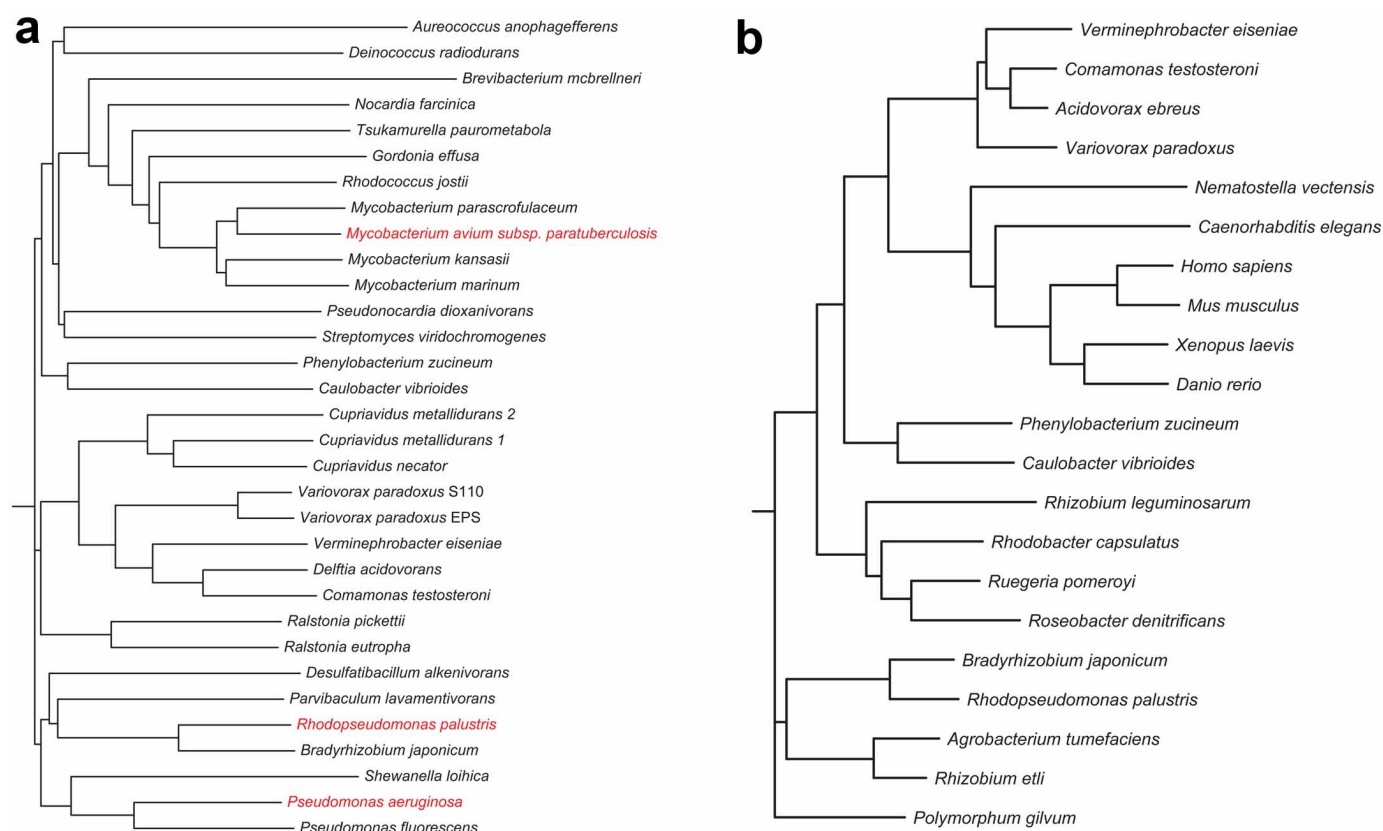
Extended Data Figure 2 | Phylogenetic trees for selected biotin-dependent carboxylases. a, Phylogenetic tree for LCC homologues in a collection of organisms. The three homologues studied in this paper are shown in red.

b, Phylogenetic tree for PCC homologues in a collection of organisms, based on a sequence alignment of the β subunit. Modified from an output from the Phylogeny.fr server[44].

BC ▶

MapLCC		
MapLCC	1	...MTAS**LLV**ANRGEIA**LRII**RTATE**LGM**RT**VAV**YAADDAHSPHVHAADEAMPLPGSGPPAYL**NQAALLAVA**ENTG**AT**L**IHPGYGFLS**EN**A**EF
RpLCC	1	..MPFRK**LLI**ANRGEIA**IRI**ARAAESG**LTT**VA**I**HPADDAAS**LHV**RT**I**AD**HA**EQIPGRGARAYL**DIEA**VIVAA**KAA**CC**DA**L**HPGYGFLS**EN**A**DL
PaLCC	1	MTAPFNA**LLI**ANRGEIA**IRI**ARACAD**LGI**RS**VAV**FAE**DDAAS**LHV**RK**A**DV**A**LPLAGRG**VAAYL**DMDR**LVALA**LEQGCCE**AI**H**PGYGFLA**EN**G**EF

MapLCC		
MapLCC	91	AK**A**C**AAASHTFVGP**DGRV**LE**L**LGNKTAARRAA**IAAG**VP**T**L**AA**TDGPSG**VEDIE**AFF**AAHP**..GG**IMIKALAGGGGRGMRK**VHN**AAQ**IAGAYQR
RpLCC	92	A**RR**C**AEEGITPIGP**SPE**A**L**RIF**GD**KVAAKELAKRCG**VPIIA**GTIGPSTL**EEVQAFFASLG**DKPA**VMTKAMAGGGGRGMR**VVE**RA**EELADAYAR
PaLCC	94	A**RR**C**QRAGIHFVGP**QAE**VLD**L**F**GD**KAAARALA**ERLE**VP**L**VAGINRAVS**VEEAEAF**LEGLG**DGAA**VMLKALAGGGGRGMRA**VEE**VAQ**LADAYRR

MapLCC		
MapLCC	182	C**AAEARLGFGDPA**L**FAEA**L**LGDARHIEV**QVV**AAPA**GH**Q**T**HALAV**GDRDCSMQRRY**QKLIEIAP**AQGLS**EELRRALH**QA**AVR**LCAR**V**GLRGLAT
RpLCC	185	C**QSEAKAAFGYEGVA**Y**AERLIRN**A**RHIEV**QII**GDRH**GNVS**Q...**LWE**R**CTIQRRN**QKLIEIAP**SPSIS**EGLRS**R**IVD**AA**KT**LAAAANYDSLGT
PaLCC	187	C**RAEAQAAFGRDELY**V**EQRVAR**A**RHIEV**QVL**GDGS**GAVS**H...**LWE**R**DCSLQRR**Q**KLLEIAP**SPDLP**EAT**R**EA**LIDCALR**MAGA**V**RYRGIGT

MapLCC		
MapLCC	275	V**EFLVSGE**......RF**VFLEVNPRIQVEHTVTE**QTTG**LDLVAVQLATAGG**ESYYR**LGL**PAGIASDGTEVIGEPAAQRG**IAIQLRVN**AET**FGAD
RpLCC	275	F**EFLVDSEAGEGDGAF**AF**IEAN**P**RIQVEHTVTE**EVLS**VDLVRSQLAVAGG**AMLTA**LGL**DQAS..........VPRPRG**FAIQLRVN**MET**MDAK
PaLCC	277	F**EFLVDDE**...RPGRF**YFMBAN**P**RIQVEHTVTE**EVTG**VDLLHAQLRLAAG**MELAA**ALGL**ERP............PAIGG**CAVQLRIN**LET**LAAD

MapLCC		
MapLCC	362	FSV**LPSAGTLTAFCPPSGPCVRVD**TYGRRG**LVVS**PQ**YDSLLAK**VIV**AVHGSS**WRAAMRKADTALS**EFGV**EGVAT**NIG**F**L**RR**LLAEQRLETGW**V
RpLCC	358	GAT**KPTGGTLAIFEPPSGPCVRVD**TFGY**AGYTTS**AA**YDSLLAK**LVV**HGR.GD**WAEAVAKASRALR**EFRAG**GVVT**NIP**L**L**QA**I**LAHNDF**KANR**V
PaLCC	355	GSA**RPAVGTIGAYEPPSGPGLRVD**GYGYAGYRVS**PSY**DSLLAK**LVV**RG..ADY**PAALRRAYRALC**EFRL**EGVAS**NLD**L**L**RN**LLLH**PAVQANR**V

BC-BCCP linker ◀▶   ◀▶ BCCP

MapLCC		
MapLCC	455	N**TDFVDEKLPE**L**AAAALTHQPGP.............**QPGAVE**LYPGEDALR**AQLAGT**VVEV**.APEG**ADYPAGA**QLVVLE**AMKMQHVLAA**PDA
RpLCC	450	S**TSFIDRHVAD**L**VEAAQQFDQPLIATGDADNRSSAAAAASKEP**VPEGAVAIAA**PLQGT**VVAIT**VAEG**DVVRPGQQLAVIE**SMKME**HLVAAE**QG
PaLCC	446	D**TRFVESHLET**L**LAPIPASHPRLRAE....CPLAEDAAPARVE**APLGSLPLSAPSS**GVLVALE**VADG**ERVRAGQRVAILE**AMKME**FEVKA**PGG

BCCP-CT linker ◀▶   ◀▶ CT N domain

MapLCC		
MapLCC	533	LR**T**V**RVLVTPGQVVGTGDPLLVFTRT**TGGD**DGESETAAT**DLDLTRAD**L**DEVIERHARTLDEGRPDAVAKRHRQGR**RTARENI**DD**LVDPGSFVE
RpLCC	543	GRI**RRIVATDGVTLMQGEPILYFEPQEVEGDHVVAEQDVDLDAIRPDLAEMLARQANTLDENRPDSVGRRKTNQR**TARENVAQ**L**VDEGSFME
PaLCC	535	GIV**RRLAASLGEPLEEGATLLFLEPTEDDDEQAPTEQALDLAHIRADLAEVLERQAALGDERRPQALARRKTGQ**RTARENVLD**L**LDEGSFSE

MapLCC		
MapLCC	626	Y**GALAIAAQRSRRSEEDLIANTPADGLVAGLATVGADRFGPAAAQAVVA**SYDYTV**LAGTQGMRNHAKTDRVFDLAARKRLPVVLFAEGGGGRP
RpLCC	636	Y**GSLAIAAQRRRRPLDDLIKNTPADGLVTGVATVNAKQPGEHAARCMVIAYDYTVLAGTQGHMNHKKIDRMLTLVEQWRMPLVF**Y**AEGGGGRP
PaLCC	628	Y**GGFALAAQRRRRSAEELLELSPADGLVAGTGTVGAASFGVQAARCLVLAYDYTVFAGTQGVMNHKKTDRLLGLAEQWRLPLVLFAEGGGGRP

MapLCC		
MapLCC	719	GDTDVGGA**A**GLDVPTFRMLAGLRGRVPLVSVSGRCFAGNAALAGVCDVIIATPDANIGMGGPAMIEGGGLGVYPPEAIGPIAVQRRNGVVGL
RpLCC	729	GDTDRLGMTGLDGPSFVQFAKLSGLVPLVSVSGYCFAGNAAMLGCCDVIIATQNASIGMGGPAMIEGGGLGVYHPAEVGPVSFGSPNGVVDI
PaLCC	721	GDTDFVGVAGLDCHTFVGMARLSGLVPLVGVVSGRCFAGNAALLGCCDVIIATRDATIGMAGPAMIEGGGLGRFAAEEVGPTGVQGPNGVIDV

CT C domain ◀▶

|		
MapLCC		
MapLCC	812	V**ARDEAHAVSLAKQYLSYFQGRLDRWQAPDPRVARHVVPQNRLRAYDVHRAIEAIVDVGSVLELRPDYGVGIVTALVRVEGVPYGLIANSTHH
RpLCC	822	L**VQDEEEATRVAQKYLSYFQGAVABWKAPDQRLLRQAIPENRLVYDIRSVIDLIADEESVLELRRDFGVGMITAFIRIEGKPFGLIANNPKH
PaLCC	814	L**VEDEAEAVAVARRYLGYFQGRLPDWSCADQRELRHLVPENRLRAYDIRQAIEVLADRGSVLELRRQFAPGLVTALLRIEGRAFGLIANNPGH

MapLCC		
MapLCC	905	**LGGAIDAEAADKAGDFLALGESFGLPVISLCDTPGFMVGPDAETQAAVRRFGRMFVLGARLTVPLGMIILRKGYGLGAMAMAGGSFRAPQFTV
RpLCC	915	**LGGAIDADAGDKAARFLQLCDAFDIPIVSLCDTPGFMVGPEAEKTAIVRHVARMFVTGASLTVPLFGIVLRKGYGLGAQSMIGGGPHASFETA
PaLCC	907	**LGGAIDAAAGDKAARFMQLCDAFDIPIVSLCDTPGFMVGPEAEKQATVRHVSRMFVSAASLTVPFFTVVLRKGYGLGAQMAAGSFHSPLFTV

◀

MapLCC		
MapLCC	998	**AWPTGEIGGMGLESAVRLGYRKELAAVTDPAERQNLFDRLVAAAYQHGKALRAATTFELDDVIDPAASRAWITRLSGG...............
RpLCC	1008	**AWPTGEFGGMGLEGYVRLGFRKEMEAIADPAEREAYYKNKVAELYANGKAVSIASVFEIDNVIDPAETRRWVMAGLRSVPTPPARTERKRPCI
PaLCC	1000	**AWPSGEFGAMGLEGAVRLGFAKELAAEEDPQRREALFRGMVDKAYRNGKALNMASYLEIDAVIDPAETRAWLLRGLAVAGEPAPRAGRKRPFV

|		
MapLCC		
MapLCC		...
RpLCC	1101	DTW
PaLCC	1093	DTW

Extended Data Figure 3 | Sequence alignment of long-chain acyl-CoA carboxylases (LCCs) from *M. avium* subspecies *paratuberculosis* (MapLCC), *R. palustris* (RpLCC), and *P. aeruginosa* (PaLCC). The various domains in the proteins are labelled. The BCCP domain has two linkers to the rest of the protein. Modified from an output from ESPript[45].
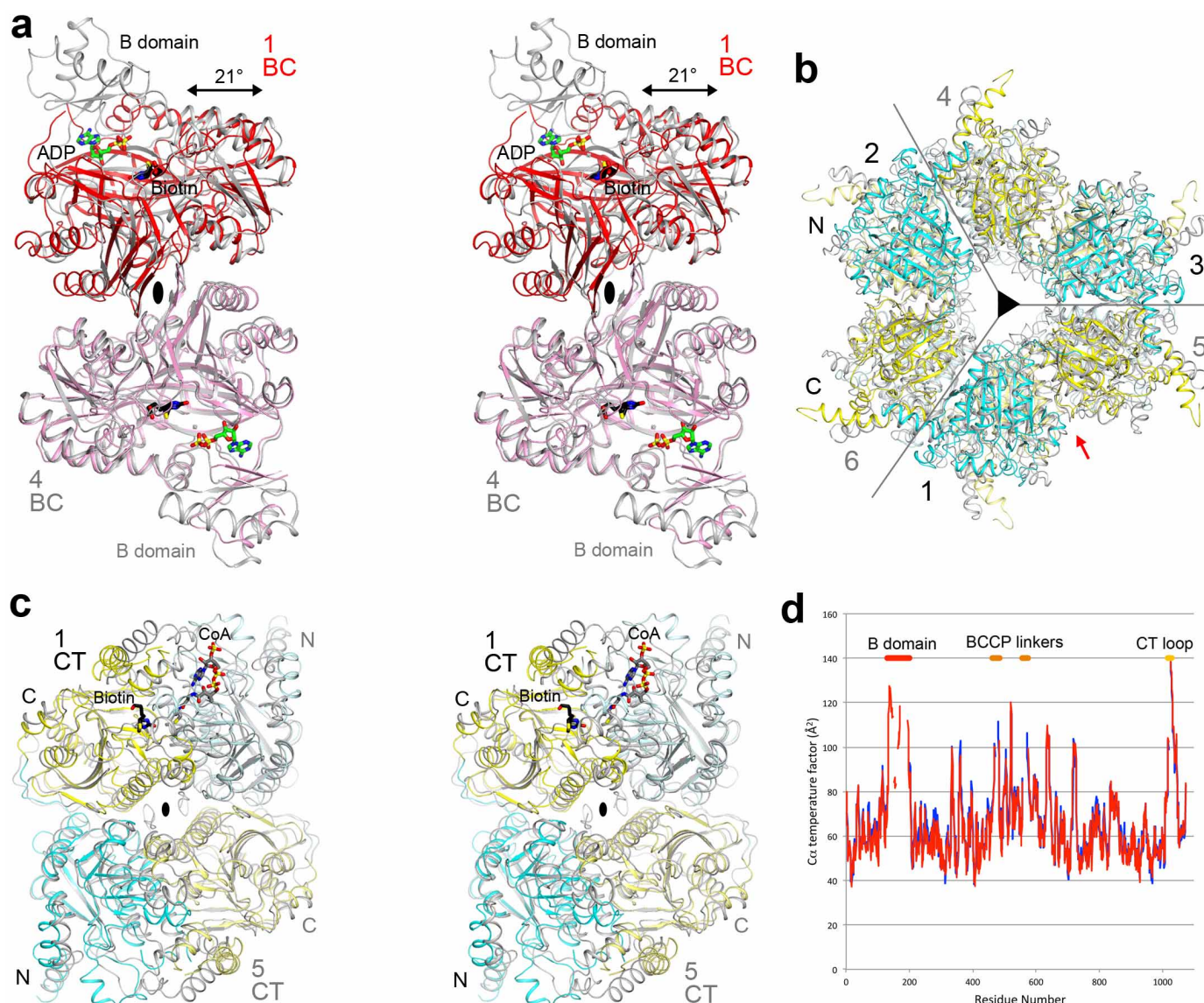
Extended Data Figure 4 | Structural comparisons of domains in LCC with related enzymes. a, Stereo drawing of the overlay of the structure of the BC domain dimer of MapLCC (in colour) with that of BC subunit dimer of *E. coli* ACC (in grey)[20]. The bound positions of biotin (black) and ADP (green) in the *E. coli* BC structure are also shown. The two-fold axis of the dimer is indicated by the black oval. With the two monomers at the bottom overlaid, a difference of 21° in the orientations of the two monomers at the top is observed. Most of the B domain of BC is ordered in one of the two monomers of MapLCC. In the other monomer, only weak electron density is observed for a few segments, and the B domain is not modelled. **b**, Overlay of the structures of the CT domain hexamer of MapLCC (in colour) and the β subunit of PCC (in grey)[10]. Each enzyme is highly conserved across species; the overlay should therefore be meaningful. **c**, Stereo drawing of the overlay of the CT domain dimer of MapLCC (in colour) and the β subunit of PCC (in grey). The view is down the red arrow in **b**. The bound position of biotin in the holoenzyme is shown in black. The position of CoA is modelled on that of CoA bound in the active site of the CT domain of yeast ACC[29]. **d**, Plot of the temperature factor value of each Cα atom in the two monomers (in red and blue). Several linker regions with high temperature factor values are indicated.
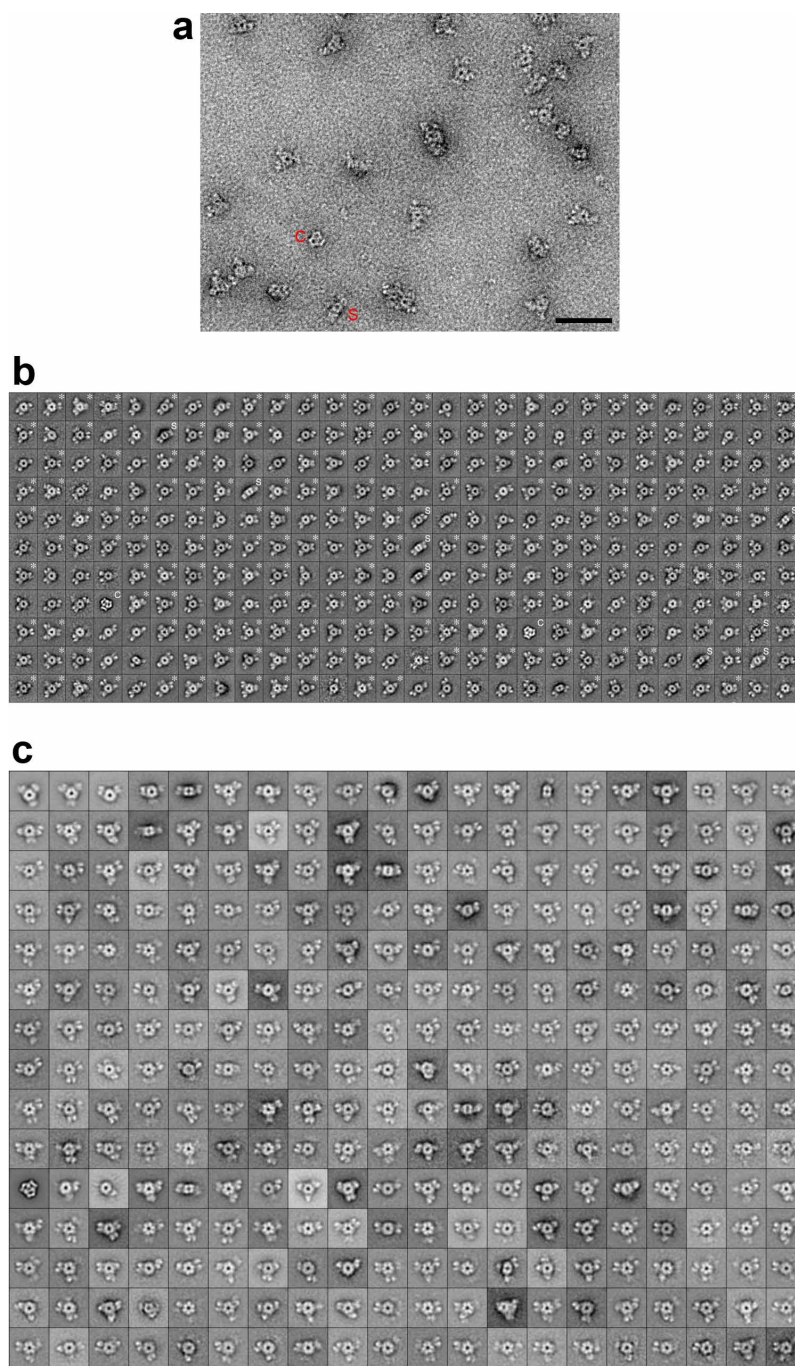
**Extended Data Figure 5 | Electron microscopy studies of LCC.**
a, Representative raw image of negatively stained MapLCC. 'S' marks a side view of the holoenzyme, and 'C' indicates a contaminant. Scale bar, 500 Å.
b, The 308 class averages of negatively stained MapLCC obtained from 19 generations of the iterative stable alignment and clustering (ISAC) procedure[40] implemented in SPARX[21]. These class averages represent 65% (15,932 particles) of the entire data set (24,535 particles). Averages representing side views are marked with 'S', averages that were used to create Supplementary Video 1 are marked with an asterisk, and averages that represent a contaminant are marked with 'C'. The side length of the individual panels is 340 Å. **c**, The averages obtained by classifying all 24,535 particles of negatively stained MapLCC into 300 classes using K-means classification in SPIDER[39]. Averages are shown in rows, with the most populous class at the top left and the least populous class at the bottom right. The side length of the individual panels is 340 Å.
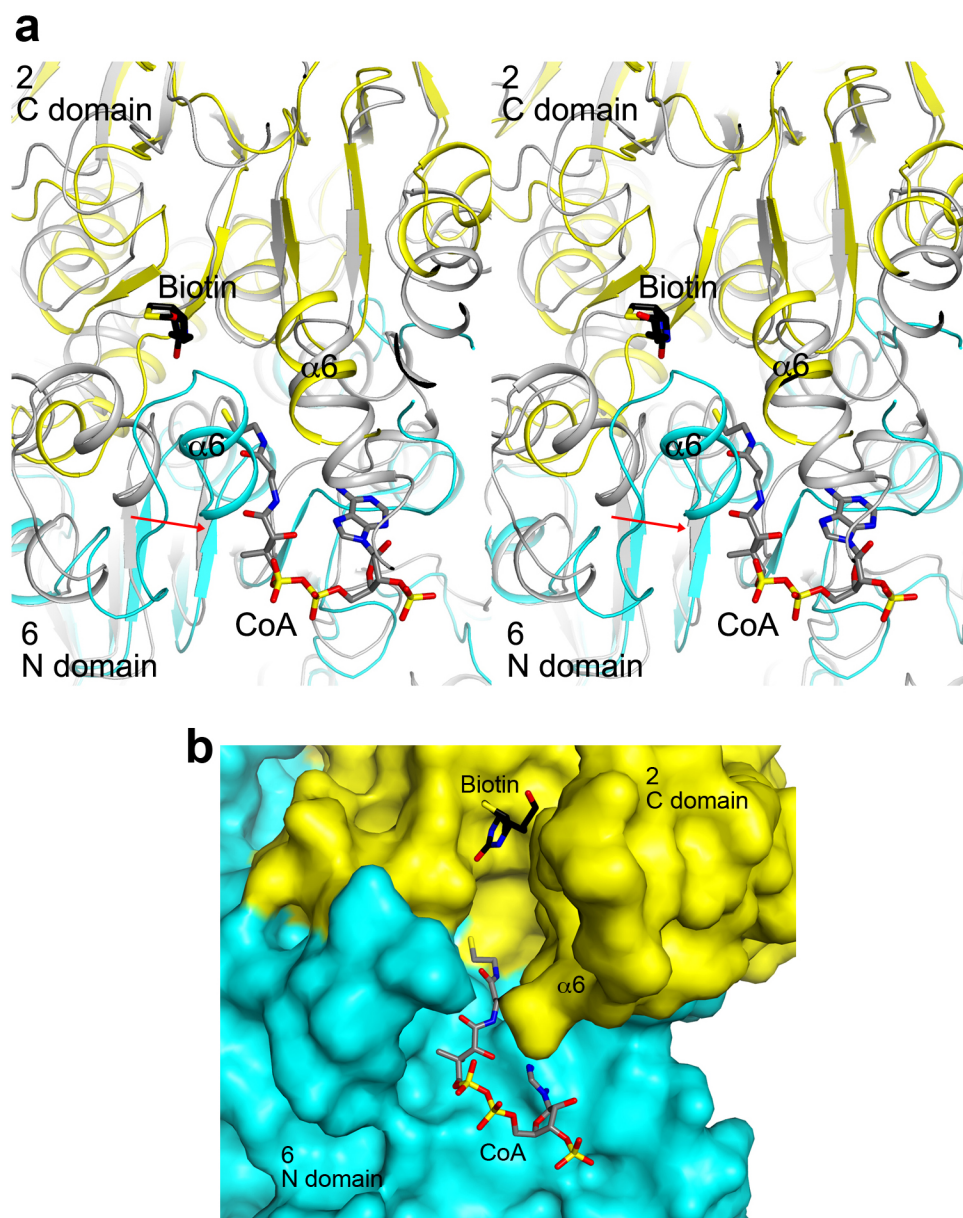
Extended Data Figure 6 | The CT active site of LCC. a, Stereo drawing of the overlay of the CT active site (cyan and yellow) of MapLCC with that of PCC (grey)[10]. The model of CoA was obtained from the structure of the complex with yeast ACC CT domain[29]. The α6 helix in the N domain of monomer 6 shows a more closed conformation (indicated by the red arrow) and clashes with the CoA model. There is also a clash with the adenine base of CoA. There may be a conformational change in this region of MapLCC for CoA binding. **b**, Molecular surface of the CT active-site region of MapLCC. The α6 helix in the N domain of monomer 6 was removed for a clearer view of the active site.
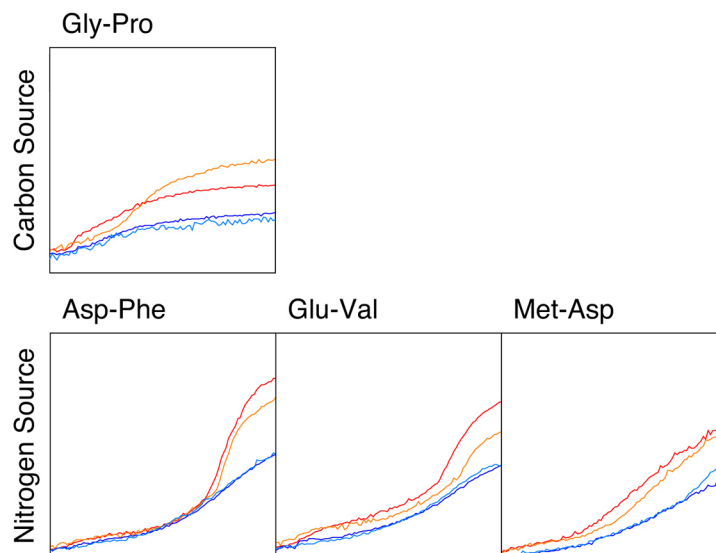
Extended Data Figure 7 | Phenotypic differences between wild-type and LCC knockout (ΔPA14_46320) *P. aeruginosa* strains, revealed by a colorimetric assay that monitors the reduction of a tetrazolium dye. The conditions were identified from a screen that sampled 1,920 different media (Biolog Inc.). Assays were performed twice in each medium for the wild-type (red and orange) and mutant (blue and cyan) strains. Shown are activity profiles for strains incubated with Gly-Pro as the sole carbon source (top panel) and Asp-Phe, Glu-Val and Met-Asp as the sole nitrogen source (bottom panels). For each panel, the horizontal axis is time (24 h), and the vertical axis is OmniLog signal[27].

Extended Data Table 1 | Kinetic parameters of *R. palustris* LCC towards various substrates

Substrate	K_m (mM)	k_{cat} (s^{-1})
Acetyl-CoA (C2)	2.0±0.3	0.77±0.05
Propionyl-CoA (C3)	2.2±0.5	0.54±0.07
Butyryl-CoA (C4)	0.82±0.16	0.35±0.03
Hexanoyl-CoA (C6)	0.20±0.04	0.30±0.01
Octanoyl-CoA (C8)	0.37±0.06	0.33±0.01
Decanoyl-CoA (C10)	0.033±0.005	0.28±0.01
Lauroyl-CoA (C12)	0.019±0.003	0.24±0.005
Myristoyl-CoA (C14)	0.026±0.002	0.45±0.01
Palmitoyl-CoA (C16)	0.0058±0.0005	0.16±0.003
3-methylcrotonyl-CoA	1.1±0.1	0.074±0.004

The errors are standard deviations from fitting one titration curve to the Michaelis–Menten equation.

Extended Data Table 2 | Data collection and refinement statistics

	MapLCC
Data collection	
Space group	$P2_13$
Cell dimensions	
a, b, c (Å)	220.9, 220.9, 220.9
α, β, γ (°)	90, 90, 90
Resolution (Å)	50-3.0 (3.1-3.0) *
R_{merge}	9.6 (44.6)
$I/\sigma I$	10.3 (1.9)
Completeness (%)	91 (72)
Redundancy	3.3 (2.1)
Refinement	
Resolution (Å)	50-3.0
No. reflections	64,953
R_{work}/ R_{free}	20.9 / 26.2
No. atoms	
Protein	14,632
Ligand/ion	0
Water	0
B-factors	
Protein	66.4
Ligand/ion	–
Water	–
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (º)	1.4

Two crystals were used for data collection.

* The highest-resolution shell is shown in parenthesis.

Extended Data Table 3 | Primers used for making the LCC deletion mutant

Primer	Sequence (5' to 3')
ΔPA14_46320 flank 1F	CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGAT GCTGCCTGCTCTACATGCT
ΔPA14_46320 flank 1R	CCTTCAACGCCTTGCTGATCCAGCTACCTGGAGATCGAC
ΔPA14_46320 flank 2F	GTCGATCTCCAGGTAGCTGGATCAGCAAGGCGTTGAAGG
ΔPA14_46320 flank 2R	GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT GGCGCGACCAGTAGAGATT

PICK UP PYTHON

A powerful programming language with huge community support.

BY JEFFREY M. PERKEL

Last month, Adina Howe took up a post at Iowa State University in Ames. Officially, she is an assistant professor of agricultural and biosystems engineering. But she works not in the greenhouse, but in front of a keyboard. Howe is a programmer, and a key part of her job is as a 'data professor' — developing curricula to teach the next generation of graduates about the mechanics and importance of scientific programming.

Howe does not have a degree in computer science, nor does she have years of formal training. She had a PhD in environmental engineering and expertise in running enzyme assays when she joined the laboratory of Titus Brown at Michigan State University in East Lansing.

Brown specializes in bioinformatics and uses computation to extract meaning from genomic data sets, and Howe had to get up to speed on the computational side. Brown's recommendation: learn Python.

Among the host of computer-programming languages that scientists might choose to pick up, Python, first released in 1991 by Dutch programmer Guido van Rossum, is an increasingly popular (and free) recommendation. It combines simple syntax, abundant online resources and a rich ecosystem of scientifically focused toolkits with a heavy emphasis on community.

HELLO, WORLD

With the explosive growth of 'big data' in disciplines such as bioinformatics, neuroscience and astronomy, programming know-how is becoming ever more crucial. Researchers who can write code in Python can deftly manage their data sets, and work much more efficiently on a whole host of research-related tasks — from crunching numbers to cleaning up, analysing and visualizing data. Whereas some programming languages, such as MATLAB and R, focus on mathematical and statistical operations, Python is a general-purpose language, along the lines of C and C++ (the languages in which much commercial software and operating systems are written). As such, it is perhaps more complicated, Brown says, but also more capable: it is amenable to everything from automating small sets of instructions, to building websites, to fully fledged applications. Jessica Hamrick, a psychology PhD student at the University of California, Berkeley, has been ▶

programming in Python since 2008 and uses it in all phases of her research. In a study investigating how people manipulate geometric objects in their minds, for instance, she used the language (as well as JavaScript) to generate different shapes, present those to study participants, record their choices and analyse the data.

Despite its general-purpose power, Python is considered less painful for beginners to learn than other options. That accessibility is a function of both the language itself and the resources that have been built up around it (see 'A Python toolkit'). For example, software execution can be interactive — type a command, get a response — whereas in C, a compilation step is required to translate the code into an executable file, which complicates the process for neophytes. The language is also generally easier to handle; users do not have to predefine whether a variable will hold numbers or text, for instance. The classic programming exercise of printing 'Hello, world!' to the screen is as simple as it can be in Python — just type `print("Hello, world!")` at a Python prompt and hit Enter. "It's easier to teach novice programmers how to get things done in Python than in C++ or C," says Brown, now at the University of California, Davis. Python is in fact a popular choice for introductory programming classes in general.

The community aspect is particularly important to Python's growing adoption. Programming languages are popular only if new people are learning them and using them in diverse contexts, says Jessica McKellar, a software-engineering manager at the file-storage service Dropbox and a director of the Python Software Foundation, the non-profit organization that promotes and advances the language. That kind of use sets up a "virtuous cycle", McKellar says: new users extend the language into new areas, which in turn attracts still more users.

The community seems especially dedicated to encouraging women, Brown notes. There are numerous women-centric resources available, including workshops offered by the Hackbright Academy in San Francisco, the non-profit organization Ladies Learning Code in Toronto, Canada, and the global mentorship group PyLadies. As a master's student at McGill University in Montreal, Canada, Emily Irvine picked up Python to help her make sense of neuronal electrophysiology data. She was attracted to the language because of its "simple syntax" and "massive amount of online support". But just as important was the wider Python community, says Irvine, who will start a PhD in neuroscience at Dartmouth College in Hanover, New Hampshire, this autumn. At the PyCon conference last April in Montreal, "they just had such a welcoming atmosphere, especially towards women and scientists".

Educational resources also abound. The Software Carpentry Foundation runs a series of two-day workshops that focus on scientific programming, and many of its educational resources are available online. Online classes

---

A PYTHON TOOLKIT

How to get started

- Install Python through Anaconda or Enthought Canopy and find documentation at the Python Software Foundation
- Lessons for beginners can be found at Software Carpentry; Learn Python the Hard Way; Codecademy; and Think Python
- Other online resources on Python programming include a course from the Massachusetts Institute of Technology in Cambridge, lecture notes from Thomas Robitaille at the Max Planck Institute for Astronomy in Heidelberg, Germany, and a widely recommended essay from Google's head of research, Peter Norvig
- Open-source packages are available through SciPy.org
- Guides to programming and community support are available through Ladies Learning Code and Stack Overflow. PyCon.org lists conferences around the world.

Links to these resources can be found at go.nature.com/x2pzh1

---

are also available through Coursera in Mountain View, California, and Edx in Cambridge, Massachusetts, as are do-it-yourself tutorials, such as those hosted by Codecademy in New York City. (Because Python is named in honour of Monty Python, these tutorials often work references to the British comedy troupe into their exercises: one Codecademy exercise, for example, is to capitalize and calculate the length of the phrase 'the ministry of silly walks'.)

Irvine taught herself to code using online courses and a healthy dose of the programming Q&A site stackoverflow.com. Today, she says, she considers herself somewhere between a beginner and an intermediate Python programmer, or 'pythonista', as they are sometimes called.

THE FULL MONTY

Of course, user-friendliness is meaningless if researchers cannot write the software they need. That is where Python's packages, which extend the language with new functionality, come into play. "Python was developed as a language with a philosophy that it was 'batteries included'," McKellar says — it has built-in capabilities that make it easy to get started right out of the box. But, "it also has a very mature package ecosystem around it. Anything that you could possibly write code to solve, people have

NATURE.COM
For more on scientific software, apps and online tools, visit:
nature.com/toolbox

---

written libraries to make that easier for you."

Scientific programmers, irrespective of their discipline, routinely use a small set of core packages: NumPy (mathematical arrays), SciPy (linear algebra, differential equations, signal processing and more), SymPy (symbolic mathematics), matplotlib (graph plotting) and Pandas (data analysis). Another popular tool, Cython, addresses Python's relatively slow execution speed. Cython optimizes certain aspects of Python code, such as 'for' loops (used to instruct a program to repeatedly run a specific block of code) that are notoriously slow, essentially by converting them into C. "You can get speed-ups that are up to 1,000 times faster than standard Python," says Paul Nation, a theoretical physicist at Korea University in Seoul.

The IPython Notebook is another popular package — Howe terms it "a coder's lab notebook" — that allows users to interleave data, code and explanatory text in a single browser-based page, rather than in separate files (see *Nature* **515**, 151–152; 2014).

Beyond the core packages, software packages exist for just about every scientific discipline, including scikit-Learn for machine learning, Biopython for bioinformatics, PsychoPy for psychology and neuroscience and Astropy for astronomers. Thomas Robitaille, a coordinator of the Astropy project and a researcher at the Max Planck Institute for Astronomy in Heidelberg, Germany, says that Astropy was created to reduce duplicated effort between research groups. It gives users a core set of abilities, such as ways to convert coordinates from one astronomical mapping system to another, and a unified interface for reading and writing different data file formats, manipulating images and carrying out cosmological calculations. QuTip, another Python package, enables researchers working on quantum mechanics to define a system and then simulate how it behaves. The project was launched in 2010 by Nation and Robert Johansson, a postdoctoral fellow in RIKEN's Interdisciplinary Theoretical Science Research Group in Wako, Japan, to adapt into Python a MATLAB package that Nation was using.

Such packages are key enablers of McKellar's 'virtuous cycle'. But researchers could probably do their work using any language, provided they put in the time to learn it. (Indeed, in many languages, including Python, it is possible to run algorithms written in a different language, thereby allowing researchers to reuse their old code.) The difficult part of learning to program lies with the fundamentals, says Brown — once a researcher has those nailed down, adapting to a new language is just a matter of syntax. What matters most in the early stages is having a good support network. "Pick the programming language based on what people around you are using," Brown advises. Increasingly, that language is Python. ∎

Jeffrey M. Perkel *is a writer based in Pocatello, Idaho.*

CAREERS

WAVEBREAKMEDIA/SHUTTERSTOCK

Fresh perspective

Undergraduate researchers can boost a lab's energy and work, but need help to flourish.

BY PAUL SMAGLIK

Alan Berkowitz was taken aback when a colleague told him that undergraduate research is an oxymoron. Berkowitz, head of education at the Cary Institute of Ecosystem Studies in Millbrook, New York, had reason to be surprised. Since 1988, he has been running an undergraduate research programme at Cary, and he recruits up to a dozen students a year for the institute's 12-week summer session. He says that undergraduates are much more than cheap labour and can contribute to further insight. In training them, he says, he has seen his own scientific thinking sharpen.

Had Tracy Johnson not gained research experience as an undergraduate, she says, she probably would not have done a science PhD. "As with many students who love science, I'd only considered medical school," says Johnson, now a molecular biologist at the University of California, Los Angeles. The experience changed her career trajectory. "When I was in the lab, it was like the world opened up. I understood what the process was. I learned you could create new knowledge. If you had the right intellectual tools, you could ask and answer questions." Her adviser helped her to realize that she could contribute to science, even at this early point in her career. "It was a wonderful experience because he was a terrific mentor, and he created a research environment that was rigorous but fun," she says. "The postdocs and graduate students in that lab seemed to have fun working together and doing great science. That experience set the bar for what I wanted my own research lab to be like."

Whereas the benefits of undergraduate research for the student might seem obvious, some — like Berkowitz's colleague — wonder why a principal investigator (PI) would ever want to staff their lab with inexperienced people who have busy coursework schedules that make it hard to attend regular meetings or get into a work rhythm. Rotating undergraduates into and out of a lab every summer or semester means that PIs must find projects that do not require a long-term commitment. And, of course, undergraduates will have a steep learning curve just to master the basic language of the lab, let alone its protocols and techniques.

A PI who decides to bring in undergraduates will need to plan for their inexperience and time constraints. Experiments cannot be left unattended with the assumption that an undergraduate will know what to do. It may be necessary, say veteran lab heads who have hosted many undergraduates, to be more patient than with older trainees, and to spend more time in the lab rather than analysing data or writing grant applications. But the pay-offs can be significant. Some of those benefits include providing mentoring opportunities and soft-skill building for graduate students and postdocs. And sometimes an undergraduate's unbiased opinion can benefit a research project. At other times, simply having more hands in a lab can speed up the work. Furthermore, a PI and other lab members can help to drive a young student's career choice, just as they did with Johnson's.

PROJECT MANAGEMENT

When undergraduates form part of a lab team, it is important to come up with experiments that have a definitive end point and can be divided into small, manageable blocks, says David Asai, who runs the undergraduate research programme at the Howard Hughes Medical Institute in Chevy Chase, Maryland, which places students with investigators around the nation. "Undergrads are very busy. They have classes they have to take. They are involved in lots of other things that are important," he says.

Susan Singer, director of the US National Science Foundation's Research Experience for Undergraduates programme, says that she worked with undergraduates for some 30 years in her developmental-biology lab. She agrees that it is crucial to structure research projects to align with students' availability and experience. During one gene-expression experiment, for example, she parcelled her students into groups, each of which looked at a different gene. Every four hours, different students in each group would collect plant tissue samples and RNA, ensuring that no person was solely responsible for any one data point, and no important information ever went missing. "You have to build in redundancies, checks and balances," she says. ▶

Jackie Tanaka, who leads an undergraduate research programme at Temple University in Philadelphia, Pennsylvania, finds that investing more time up front pays off. Research projects typically require a lot of repetitive tasks that could elicit complaints of boredom if the students do not understand why the jobs need to be done. "Students have to be trained in the need for care and reproducibility," she says. "They don't necessarily realize the importance of what they are doing." Working with undergraduates is also likely to require tact and diplomacy, she says. They sometimes have difficulty accepting feedback, especially when it is constructive rather than positive. "It takes patience," she says.

However, undergraduates often have greater patience for repetitive tasks than do more-senior scientists, says Ritwick Sawarkar, a group leader at the Max Planck Institute of Immunobiology and Epigenetics in Freiberg, Germany. "The younger people are much more cheerful and bring in so much excitement to the lab," he says. "Every DNA gel brings them joy."

VALUABLE CONTRIBUTIONS

Sometimes, that naivety even translates into scientific success. During one meeting, Sawarkar's group was stuck trying to find a way to inhibit a protein in a cell's nucleus. An undergraduate with a chemistry background offered a suggestion that the group had never considered. Sure enough, when the researchers checked the chemistry literature, they found that her suggestion was a viable method that they eventually used. "She opened our eyes to look into an area we wouldn't have normally considered," says Sawarkar.

Getting students comfortable enough to pitch in takes effort, however. Catherine Drennan, a structural biologist at the Massachusetts Institute of Technology in Cambridge, says that she starts by having her postdocs and graduate students teach her undergraduates basic protein-crystallization techniques. She then assigns the undergraduates proteins to crystallize. Eventually, the students pick their own proteins and crystallization methods. "My overall goal is to train them to do basic stuff," she says. "Once they learn the ropes they can carve out their own puzzle."

Her lab lends itself well to undergraduate research because protein crystallization requires short efforts over a long period. Students can put a protein into solution, leave it for a few hours, perhaps while attending a class, then return to check on it. They then tweak the solution by changing the concentration, temperature and pH level, among other factors, until they get the combination right. "Sometimes the first thing you try works and sometimes you have to try hundreds," Drennan says. Having undergraduates take on this step frees her graduate students and postdocs to characterize the crystallized proteins. She has also seen undergraduates succeed where older trainees have failed, thanks to their persistence. "They are willing to try to just problem-solve," she says. "They can help to rule out a whole bunch of things that don't work."

Martin McLaughlin loves working in Drennan's lab, where he started in his first year at university. His experience there has given him the confidence to pursue a scientific career after he graduates this year. "It's a very close-knit environment," he says. "You can walk around the lab and ask anyone a question. We take care of each other." First-year student Devany West, who joined the lab just last month, says that she especially values the chance to find out what it is like to be a researcher while the stakes are lower because neither her degree nor career depend on the work. "You're being nurtured," she says. "There's an element of being intimidated. But you are not expected to be perfect. You're expected to mess up."

Undergraduates in a lab can also help PIs and older trainees to learn how to promote their own research, says Berkowitz. Having to break down research problems and explain them to an undergraduate forces the senior researcher to think about how best to develop hypotheses and design the most effective experiments to test them, he says.

In addition to the pay-off for the PI and senior lab members, there is altruistic value to hosting undergraduates. For one thing, it provides an early opportunity for students to discover that they hate the bench, before committing to doctoral programmes and postdoctoral research. "They work in the lab and confirm they love science — but they find they don't like lab work," says Singer. She has hosted undergraduates who have become researchers and those who have taken other paths, such as science journalism or lab management. "Both are successful outcomes," she says.

Micky Einstein, a doctoral student in neuroscience at the University of California, Los Angeles, credits his experience as an undergraduate researcher for giving him early insight into lab management. For instance, he knows that undergraduates' motivations for joining a lab vary. Therefore, he can help to weed out applicants who want only to add a line to their CV and are not that interested in the actual research experience.

Ramesh Pillai, a group leader at the European Molecular Biology Laboratory (EMBL) in Grenoble, France, says that managing undergraduates provides postdocs and graduate students with invaluable experience. "This is an essential skill set that you cannot sit in a classroom and learn."
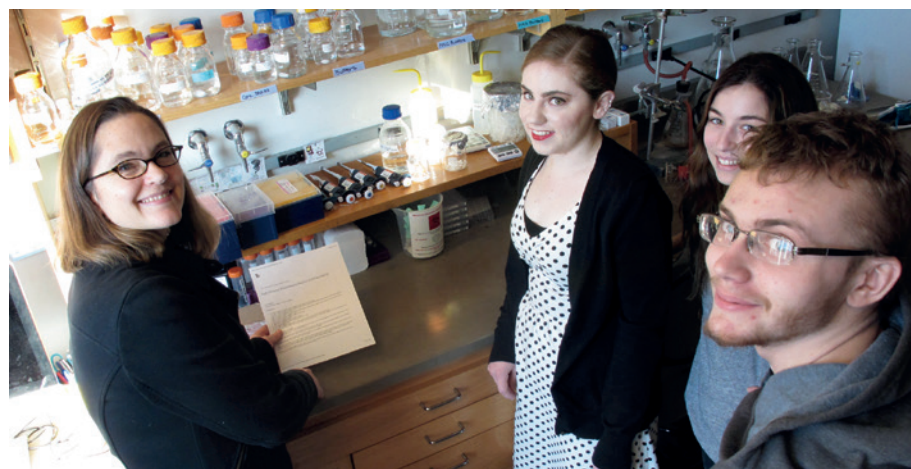
Lionel Newton benefited from postdoc and graduate mentors as an undergraduate at the EMBL in Heidelberg in 2008. Now, as a postdoc and manager there himself, he knows the importance of finding out what undergraduates already know, both practically and theoretically. They can become annoyed if a mentor repeatedly explains things they already know, but can equally get stressed if the supervisor simply hands them a piece of equipment and disappears, Newton says. "It's only through communications in the early days that you can avoid these kinds of frustrations," he says.

With good management and communication, hosting undergraduates may well be "transformative" for both them and their mentors, says Johnson. That is a much more apt way to describe undergraduate research than "oxymoron". ∎

Paul Smaglik *is a freelance writer in Milwaukee, Wisconsin.*

> *"This is an essential skill set that you cannot sit in a classroom and learn."*

Catherine Drennan (left) is keen to have undergraduates help with the research in her lab.

CORRECTION

The Careers Feature 'Speak up for science' (*Nature* **517,** 231–233; 2015) neglected to include the UK-based charity Sense About Science as an organizer of the John Maddox Prize.

Jackie Tanaka, who leads an undergraduate research programme at Temple University in Philadelphia, Pennsylvania, finds that investing more time up front pays off. Research projects typically require a lot of repetitive tasks that could elicit complaints of boredom if the students do not understand why the jobs need to be done. "Students have to be trained in the need for care and reproducibility," she says. "They don't necessarily realize the importance of what they are doing." Working with undergraduates is also likely to require tact and diplomacy, she says. They sometimes have difficulty accepting feedback, especially when it is constructive rather than positive. "It takes patience," she says.

However, undergraduates often have greater patience for repetitive tasks than do more-senior scientists, says Ritwick Sawarkar, a group leader at the Max Planck Institute of Immunobiology and Epigenetics in Freiberg, Germany. "The younger people are much more cheerful and bring in so much excitement to the lab," he says. "Every DNA gel brings them joy."

VALUABLE CONTRIBUTIONS

Sometimes, that naivety even translates into scientific success. During one meeting, Sawarkar's group was stuck trying to find a way to inhibit a protein in a cell's nucleus. An undergraduate with a chemistry background offered a suggestion that the group had never considered. Sure enough, when the researchers checked the chemistry literature, they found that her suggestion was a viable method that they eventually used. "She opened our eyes to look into an area we wouldn't have normally considered," says Sawarkar.

Getting students comfortable enough to pitch in takes effort, however. Catherine Drennan, a structural biologist at the Massachusetts Institute of Technology in Cambridge, says that she starts by having her postdocs and graduate students teach her undergraduates basic protein-crystallization techniques. She then assigns the undergraduates proteins to crystallize. Eventually, the students pick their own proteins and crystallization methods. "My overall goal is to train them to do basic stuff," she says. "Once they learn the ropes they can carve out their own puzzle."

Her lab lends itself well to undergraduate research because protein crystallization requires short efforts over a long period. Students can put a protein into solution, leave it for a few hours, perhaps while attending a class, then return to check on it. They then tweak the solution by changing the concentration, temperature and pH level, among other factors, until they get the combination right. "Sometimes the first thing you try works and sometimes you have to try hundreds," Drennan says. Having undergraduates take on this step frees her graduate students and postdocs to characterize the crystallized proteins. She has also seen undergraduates succeed where older trainees have failed, thanks to their persistence. "They are willing to try to just problem-solve," she says. "They can help to rule out a whole bunch of things that don't work."

Martin McLaughlin loves working in Drennan's lab, where he started in his first year at university. His experience there has given him the confidence to pursue a scientific career after he graduates this year. "It's a very close-knit environment," he says. "You can walk around the lab and ask anyone a question. We take care of each other." First-year student Devany West, who joined the lab just last month, says that she especially values the chance to find out what it is like to be a researcher while the stakes are lower because neither her degree nor career depend on the work. "You're being nurtured," she says. "There's an element of being intimidated. But you are not expected to be perfect. You're expected to mess up."

Undergraduates in a lab can also help PIs and older trainees to learn how to promote their own research, says Berkowitz. Having to break

"This is an essential skill set that you cannot sit in a classroom and learn."

down research problems and explain them to an undergraduate forces the senior researcher to think about how best to develop hypotheses and design the most effective experiments to test them, he says.

In addition to the pay-off for the PI and senior lab members, there is altruistic value to hosting undergraduates. For one thing, it provides an early opportunity for students to discover that they hate the bench, before committing to doctoral programmes and postdoctoral research. "They work in the lab and confirm they love science — but they find they don't like lab work," says Singer. She has hosted undergraduates who have become researchers and those who have taken other paths, such as science journalism or lab management. "Both are successful outcomes," she says.

Micky Einstein, a doctoral student in neuroscience at the University of California, Los Angeles, credits his experience as an undergraduate researcher for giving him early insight into lab management. For instance, he knows that undergraduates' motivations for joining a lab vary. Therefore, he can help to weed out applicants who want only to add a line to their CV and are not that interested in the actual research experience.

Ramesh Pillai, a group leader at the European Molecular Biology Laboratory (EMBL) in Grenoble, France, says that managing undergraduates provides postdocs and graduate students with invaluable experience. "This is an essential skill set that you cannot sit in a classroom and learn."

Lionel Newton benefited from postdoc and graduate mentors as an undergraduate at the EMBL in Heidelberg in 2008. Now, as a postdoc and manager there himself, he knows the importance of finding out what undergraduates already know, both practically and theoretically. They can become annoyed if a mentor repeatedly explains things they already know, but can equally get stressed if the supervisor simply hands them a piece of equipment and disappears, Newton says. "It's only through communications in the early days that you can avoid these kinds of frustrations," he says.

With good management and communication, hosting undergraduates may well be "transformative" for both them and their mentors, says Johnson. That is a much more apt way to describe undergraduate research than "oxymoron". ■

Paul Smaglik *is a freelance writer in Milwaukee, Wisconsin.*

Catherine Drennan (left) is keen to have undergraduates help with the research in her lab.

LAURA DOUGHTY.

CORRECTION
The Careers Feature 'Speak up for science' (*Nature* **517**, 231–233; 2015) neglected to include the UK-based charity Sense About Science as an organizer of the John Maddox Prize.

TURNING POINT
William S. Horton

Cognitive psychologist William S. Horton studies language at Northwestern University in Evanston, Illinois. But last October, he did something unusual — he co-authored a paper that had failed to replicate some of his earlier results. He explains that it was a tough decision, but has had a positive outcome.

What are your research interests?
As a graduate student, I worked on language use in conversations with a researcher who was investigating whether effective communication requires consideration of shared information. He found evidence that people are egocentric and initially give more weight to their own knowledge in conversations with others, and make language adjustments only later on the basis of feedback. As a postdoc, I had worked with someone who believed that we always keep track of 'common ground' or shared knowledge and are not egocentric initially. I developed a model that bridges both perspectives.

How did your research evolve?
I started to look at the role of memory in how people establish common ground in conversations. I showed that common ground need not be a conversation goal because other people can function as cues to retrieve relevant experiences from memory (W. S. Horton *Lang. Cogn. Process* **22,** 1114–1139; 2007).

Was it controversial?
There wasn't a strong reaction one way or another. Sarah Brown-Schmidt, a psychologist at the University of Illinois at Urbana-Champaign, wanted to build on my memory theory in her research. She recreated the experiment, but it did not replicate my results. I gave her my materials so that she could try again. That experiment failed, too, and she asked me to be a co-author in a failure-to-replicate paper.

Were you worried about doing so?
The main con was putting my name on a publication that called my earlier work into question. On top of that, I was concerned about how I would talk about this result and what it would mean for my career. The decision would have been much harder had I not yet had tenure. The pros were that it was the right thing to do and that I would be able to help to put the finding in context. Sarah and I both had a sincere interest in making clear that although this study didn't replicate my results, the idea still has worth.

What did the failure-to-replicate study find?
We found no evidence that memories established in the context of other individuals helps in the recognition of shared information during subsequent interactions.

You have had a positive response to the publication. Was that surprising?
Yes. The journal I originally published in chose not to review it, so we went to *PLoS ONE* (S. Brown-Schmidt and W. S. Horton *PLoS ONE* **9,** e109035; 2014) which encourages the publication of negative results. The study got picked up on Twitter, Reddit and CBC radio. I was surprised that others found it so noteworthy.

Do you think that more researchers should publish the findings of replication attempts?
There is an increasing effort, at least in psychology, to document replications in open-source databases, such as the Reproducibility Project. Some top-tier psychology journals have adopted pre-registration reporting, in which the methods and data-analysis plans are reviewed before replication is attempted, to smooth the review process. To what extent the original author is part of that process is pretty open.

Will you try again to validate your theory?
I still very much believe in it and have other results that support it, but I may look for new ways to address the same questions.

Where do you go from here?
I'm interested in seeing how this paper gets cited. I believe in the accumulation of findings. Not every result is going to hold up. That's just how science works. ∎

INTERVIEW BY VIRGINIA GEWIN

GENDER BIAS
Seeing is not believing

Clear demonstrations of gender bias may not be enough to change attitudes. Researchers examined hundreds of online responses to reports of a study that showed experimental evidence of gender bias in science faculty members. Comments that either justified bias or denied its existence were three times more likely to come from men than from women (C. A. Moss-Racusin, A. K. Molenda and C. R. Cramer *Psychol Women. Q.* http://doi.org/zqn; 2015). Initiatives to combat gender and other bias will need to do more than offer proof that it exists, says lead investigator Corinne Moss-Racusin at Skidmore College in Saratoga Springs, New York. "We need to understand whether people are open to that evidence."

PHD TRAJECTORIES
Data wanted

A report from the US Council of Graduate Schools (CGS) in Washington DC calls for graduate schools to collect data on the careers of their PhD graduates. Such information is essential to shape programmes to help graduates to establish fulfilling careers, yet only one-third of institutions collect such data formally, concludes *Understanding PhD Career Pathways for Program Improvement*. Specifically, institutions should cooperate to develop standards and methods to track alumni careers. The publication comes at a time of growing concern about job prospects. "We hope it will be a launching pad for some real action," says CGS director of research Jeff Allum.

DOCTORAL PROGRAMMES
Online self-help

The European University Association (EUA) in Brussels has released a prototype of an online self-assessment tool for institutions with doctoral programmes. The aim is to help university leaders to decide how best to engage the international research community. Built with input from dozens of institutions, the tool can support cross-institutional discussions on strategies to build cross-country collaborations or boost international opportunities, says EUA's Thomas Jørgensen. For instance, programmes hoping to recruit more international students could be prompted to first assess their capacity for handling visas. A final version should be available in September.

HOW TO CONFIGURE YOUR QUANTUM DISAMBIGUATOR

Follow these instructions carefully — your universe depends on it.

BY STEWART C BAKER

It has come to our attention that a plurality of users has significant problems during the quantum disambiguator configuration process. These problems — many of which come from not pushing the red button located on the inside of your device — may include but are not limited to:

● Superposed instances of identical disambiguated worlds;

● Accidental creation of evil twins;

● Dead cats that are still alive (or vice versa);

● Sudden irrational activity that endangers the user's personal health as defined in a classical state (e.g. an avoidance of red buttons);

● Accidental auto-decapitation and/or persistent headaches;

● Visual hallucinations that suggest pushing the red button (if you have pushed the red button, please read the document titled *So You've Accidentally Sentenced Every Sentient Being in the Known Universe to a Horrible and Instantaneous Death* at your earliest convenience for instructions on how to revert to a pre-button world. If you have not pushed the red button, please do so at this time);

● Europe suddenly ceasing to exist, or being replaced by an improbably large banana;

● Sentient mathematical formulae which argue that the only way to really be safe from evil twins is to push the red button, no matter how compelling their evidence.

As a result of these and other problems, we would like to take this opportunity to provide our users with clear, straightforward instructions on how they may properly configure the quantum disambiguator to successfully untangle their hopelessly confused worlds.

1. Before beginning, wipe all currently disambiguated worlds from the disambiguator, being sure to ignore voices that encourage pushing the red button.

a) Push the red button.

2. Run your disambiguator through the default start-up procedure as outlined in the document titled *World-Splitting Without the Headaches: Warming Up Your New Quantum Disambiguator*.

a) If headaches persist, run through this step again, but wear a 5-star CRASH-rated helmet or duck a little earlier than you think is necessary.

3. Once you reach the configuration screen, use the following settings:

Collapse Threshold: 0.05e
State Probability Threshold: 99%
Bounding Conditions: Follow directions in document titled *From Big Bang to Heat Death: Staying in Bounds with your Quantum Disambiguator*
Evil Twin Goatee Style: Slick
Schrödinger Constant: Variable
Colour of Red Button: Blue

4. Hit 'Save'.

5. Restart the disambiguator by pushing the blue button.

After running through these simple steps, almost all users report finding themselves in a world wherein their disambiguator is running without problems. Users who still have trouble, or who are unable to find one or more of the above configuration settings on the configuration screen, may wish to consider the very real possibility that they have entered an aberrant world-instance or been manipulated by an evil twin. These users may wish to read the appendix included at the end of this document or to call or e-mail our help desk (hours vary until observed).

If, however, these steps do not resolve your problems, you may find yourself becoming increasingly frustrated. You may even consider pushing the red button, which is quite shiny and attractive and which you should probably just go ahead and push, as, statistically speaking, you've already pushed it in some other world and the worst has already happened.

No matter how frustrated you become, please do not push the red button. Doing so will set the number of potential observers in the Universe to zero, resulting in a new vacuum state across all possible worlds and causing instant death for all sentient beings, including the user. Note that if you have pushed the red button and are not yet dead, it is due to failsafes that have shunted you and your disambiguator to a pocket universe that will last just long enough for you to read the document mentioned at the beginning of this file and to regress to a pre-button-pushing world.

In certain emergencies, collapse to a new vacuum state may seem desirable (e.g. if an evil twin is about to commandeer one's body through a nefarious and highly improbable string of events involving bananas and expertly timed visual hallucinations). Even in these cases, our development team suggests first waiting until the automated nightly recalibration in the hope that your twin will be noticed by our data-checking algorithms and returned to his or her own world.

Note that if your evil twin comes from a world in which the pressing of the red button has caused the Universe to collapse to a new vacuum state, he or she will experience a horrible and instantaneous death. This is not your fault, and any feelings of guilt should be assuaged by reading the pamphlet titled *So You've Sentenced Your Evil Twin to a Horrible and Instantaneous Death*. Please take care not to mistake this pamphlet for the similarly titled *So You've Decided to Sentence Your Non-Evil Twin to a Horrible and Instantaneous Death — Again*, unless you have first pushed the red button.

Push the red button. Please do not push the red button. Push the red button. ∎

ILLUSTRATION BY JACEY